DOCUMENT RESUME

ED 245 696 IR 050 768

AUTHOR	Westerman, Mel
TITLE	Nonbibliographic Machine-Readable Data Bases in ARL Libraries. SPEC Kit 105.
INSTITUTION	Association of Research Libraries, Washington, D.C. Office of Management Studies.
PUB DATE	Jun 84
NOTE	105p.
AVAILABLE FROM	Association of Research Libraries, Systems and Procedures Exchange Center, 1527 New Hampshire Ave., NW, Washington, DC 20036 (individual kits, $15.00 prepaid).
PUB TYPE	Collected Works - General (020) -- Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Academic Libraries; *Databases; Library Administration; *Library Automation; Library Guides; Library Surveys; *Online Systems; *Research Libraries; *Use Studies
IDENTIFIERS	Association of Research Libraries; *Nonbibliographic Databases; Vendors

ABSTRACT

This document is one of ten kits distributed annually by the Systems and Procedures Exchange Center (SPEC), a clearinghouse operated by the Association of Research Libraries, Office of Management Studies (ARL/OMS) that provides a central source of timely information and materials on the management and operations of large academic and research libraries. The document presents the results of a 1984 SPEC survey on the use of nonbibliographic machine-readable databases in 34 ARL libraries along with related primary source documents. The most common access to nonbibliographic databases is through online vendors: 31 of the 34 respondents mentioned this as a source of machine-readable data. Twelve libraries acquire their own databases, but only four have collection development policies for this area. All 12 have social sciences databases; 4 have science databases and 3 include humanities databases. Five libraries handle or store their own databases; 10 share responsibility for the databases with computer centers or other campus units. More than half of the respondents went beyond offering referrals or online retrieval from major vendors. In addition to the SPEC survey results, this kit contains two committee and task force reports, two examples of manuals and descriptions, two reports on the use of nonbibliographic databases, and a list of selected resources. (DMC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED245696

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Maxine Sitts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

#105, June 1984

OMS

Systems and Procedures Exchange Center

ISSN 1060 3574

NONBIBLIOGRAPHIC MACHINE-READABLE DATABASES IN ARL LIBRARIES

More and more information is becoming available online, in some cases without any printed counterpart. As this trend continues, research libraries are finding that some information is best kept and used electronically. Bibliographic databases contain at the most abstracts of cited items, but nonbibliographic databases allow direct use of their information. Nonbibliographic databases include both numeric and text databases and combinations of the two, for example online encyclopedias and the U.S. Census tapes. Vendors have provided access to nonbibliographic databases for some time, but local files and data tapes from other sources also are increasingly common. Whether or not they handle or service these data files, research libraries need to know where the files are available, what kinds of information are included, and how users can gain access to them.

Currently, many ARL university libraries are helping students and faculty use nonbibliographic machine-readable databases, although levels of involvement vary widely according to a SPEC survey of 34 ARL members in January 1984.

BACKGROUND. In the past, most machine-readable data was available only for use on large computers. Universities established computer service departments to operate these computers, and developed specialized service units to acquire data and organize materials to support the computerized files. Libraries for the most part did not consider machine-readable data files to be within their scope, and local development of machine-readable information as well as university acquisition of data tapes normally took place outside the traditional library setting. Currently, however, research libraries are moving to provide this information. Many researchers prefer data in electronic format, and students, brought up in an online age, expect to use computers whenever they can.

SURVEY RESULTS. The most common access to nonbibliographic databases is through online vendors: 31 of the 34 respondents to the SPEC survey mentioned this as a source of machine-readable data. Twelve libraries acquire their own databases, but only four have developed collection development policy statements for this area. All 12 have social sciences databases; four have purchased sciences databases, and three include humanities databases. Five of the libraries handle or store their own databases; ten share responsibility for the databases with computer centers or other campus units.

More than half the respondents went beyond offering referrals or online retrieval from major vendors. Fifteen libraries provide advice on availability of database guides and other explanatory resources, and nine offer help on using programs. Six have library staff specifically assigned to assist users, and one offers programming services. Twenty-seven libraries seek to make users aware of database availability by various means, including catalog access and a campus directory of resources.

ISSUES. Research libraries can have various degrees of involvement with data files. They may collect only published results of research, refer users to sources of data, acquire and catalog codebooks and other guides, purchase data files, budget for institutional membership in data

The Systems and Procedures Exchange Center (SPEC) is operated by the Association of Research Libraries, Office of Management Studies, 1527 New Hampshire Avenue, N.W., Washington, D.C. 20036. Telephone (202) 232-8656.

archives located elsewhere, provide support to process raw data, assign a full or part-time staff member to work with files and work with the computing center, or outside it. As the level of involvement increases, more questions occur. Providing proper staffing, training, guides, cataloging, and public relations for database management are some of the major concerns.

Acquiring, maintaining, and servicing data files will require added expenditures. Databases come from many different sources, and are priced from less than one hundred to many thousands of dollars each. The maintenance of machine-readable files is quite different from preservation of printed materials, but computer center or other appropriate campus personnel may share their expertise, and possibly offer space as well. Service costs depend on the level of service provided. Some minimal services probably can be absorbed in the operating budget: referrals and advice at public service points, and copying documentation for databases available elsewhere in the university. Library staff searching of databases available through vendors also might be absorbed. Fees may need to be considered for higher levels of service.

Nonbibliographic databases can present difficulties for users and librarians when the information is highly specialized. Instruction manuals also can be complex. Frequently, obtaining successful results from a search depends on understanding the method by which the data was coded. For this reason, users who need information from nonbibliographic databases and staff members who search them can benefit from training programs. Even where library staff are responsible only for referrals, an understanding of both databases and their accompanying documentation is essential.

One decision to be made on staffing this function is whether to designate a reference specialist for the area or to train all or some of the reference staff. Determining the amount of time needed for the work also is a necessary step: Data file activities can include developing and maintaining a guide to collections, liaison work with vendors or data archives, assisting with acquiring and cataloging codebooks, and publicizing databases or archives membership.

Because librarians normally have had to develop a specialization in data files when it was needed, the question of including the subject in library school curricula also has surfaced.

SPEC Kit #105. Nonbibliographic Machine-Readable Databases in ARL Libraries (June 1984, 100 pages), contains SPEC survey results, two commitee and task force reports, two examples of manuals and descriptions, two documents on the use of nonbibliographic data bases, and a list of selected resources.

SPEC kits are available mainly by subscription from: SPEC Center, Office of Management Studies, ARL, 1527 New Hampshire Ave., N.W., Washington, D.C. 20036. Individual kits are available for $15.00 prepaid. Outside the U.S. and Canada, add $5.00 per order to the prepayment for parcel post shipping. (ARL library members receive kits for $7.50).

This kit/flyer was prepared as part of the OMS Collaborative Research Writing Program by Mel Westerman, Pennsylvania State University.

USES OF SPEC KITS

The Systems and Procedures Exchange Center (SPEC) is a clearinghouse operated by the Association of Research Libraries, Office of Management Studies that provides a central source of timely information and materials on the management and operations of large academic and research libraries. It facilitates the exchange of knowledge and documents through SPEC Kits, which are distributed ten times each year to ARL members and other interested libraries. The Kits include topically-arranged groupings of unedited primary source documents – selected for their value to administrators and decision-makers – that illustrate a wide range of alternative approaches to specfic issues.

Kit documents come from general membership surveys and from selected libraries contacted directly by SPEC, and most Kits are produced within six months of surveys. The documents' value comes from their variety of ideas, methods, and solutions. They are not viewed as finished products, but rather as points of departure for a library's planning efforts and as stimulants to innovative approaches to problem-solving. As such, Kits do not present answers or prescriptions for any one library; instead they illustrate how selected ARL members are planning for or dealing with particular issues. The worth of any one Kit to a particular library will depend upon the specific topic covered and the library's stage of development in that area.

Materials are selected according to the following criteria:

- Presents an approach of potential value to administrators and decision-makers
- Timely, and dealing directly with the topic under consideration
- Probability of application of ideas or thinking to other library situations
- Illustrative of actual practice, rather than theoretical
- Understandable, readable communication

All together, the materials should provide a range of alternative approaches that complement each other, provide variety, and stimulate comparison and contrast.

Libraries can take advantage of the Kit compilations in a number of ways. Administrators can evaluate the assumptions, methods, and results of other libraries' approaches; compare and contrast them; and use the learnings in their own situations. Library staff members can use the kits as professional development and current awareness tools. Committees and task forces can use them to begin a review of current practices. And the Kits can identify other persons or places to contact for further information. Back-up files in the SPEC office also are available for loan to member libraries. In addition, SPEC will conduct on-demand surveys or analyses geared specifically for a single library.

EVALUATION

Kit Title/Number _____

1. Which uses did the library make of this Kit?

2. Please indicate how useful the Kit was for these purposes.

☐ Very Useful ☐ Quite Useful ☐ Somewhat Useful ☐ Not Useful

3. Do you have suggestions for this Kit or for future Kits?

(optional)

NAME _____

LIBRARY _____

PHONE _____

Please return this form to the SPEC Center, OMS/ARL, 1527 New Hampshire Ave., N.W., Washington, DC 20036.

1/82

ISSN 0160 3582

NONBIBLIOGRAPHIC MACHINE-READABLE DATABASES IN ARL LIBRARIES

SPEC Kit #105

JUNE 1984

TABLE OF CONTENTS

	PAGE
SPEC SURVEY RESULTS	1
COMMITTEE/TASK FORCE REPORTS	
Textual and Numeric Data Files: Primary Information Sources in Machine-Readable Form, A Report from the LAUD Ad Hoc Committee on Textual and Numeric Data Bases - Librarians Association of the University of California, submitted by the University of California at Berkeley. October 1983	5
Letter, Job Descriptions, Final Report: University Libraries Taskforce on Non-bibliographic Databases. State University of New York at Albany. February 1983.	19
DESCRIPTIONS/MANUALS	
Resource Fact Sheet: Social Science Data, Archive, Review Document: Other Program and Data Resources. Yale University. October 1979.	58
Selected Sections from Data Library Procedures Manual. University of British Columbia. April 1982.	62
USE OF DATABASES	
Nonbibliographic Databases in Texas: Availability and Use - Research Proposal Texas A&M University. November 1983.	89
Description of the Inter-University Consortium for Political and Social Research and the Use of MRDFs in the Social Sciences: an anthropologist and a librarian look at the issues. University of Florida. 1983.	93
SELECTED RESOURCES	100

spec flyer
January 20, 1984
OMS
Systems and Procedures Exchange Center

ON-DEMAND SURVEY FOR THE PENNSYLVANIA STATE UNIVERSITY LIBRARIES:

THE ADMINISTRATION OF NONBIBLIOGRAPHIC MACHINE-READABLE DATA BASES IN ARL UNIVERSITY LIBRARIES

Your institution has been selected for this survey from the membership of the Association of Research Libraries to assist The Pennsylvania State University Libraries in planning, and to collect documents and information for a SPEC Kit/Flyer scheduled for publication in mid-1984.

We are interested in learning about your library's administrative practices pertaining to the acquisition, storage and servicing of nonbibliographic machine-readable data bases. NOT of interest are document delivery services, online bibliographic systems, information-on-demand services, and library networks.

Please respond to the questions below, and return the questionnaire, plus accompanying documents, by March 1, 1984. If you have questions, please call Maxine Sitts at 202 232-8656.

NAME OF RESPONDENT _____

LIBRARY/UNIVERSITY_____

PHONE NUMBER (A.C.)_____

1. Do students and faculty at your university have access to nonbibliographic machine-readable data bases (e.g., U.S. Census Tapes)?

_34_ Yes

_0_ No
IF NO, please indicate and return questionnaire.

The Systems and Procedures Exchange Center (SPEC) is operated by the Association of Research Libraries, Office of Management Studies, 1527 New Hampshire Avenue, N.W., Washington, D.C. 20036. Telephone (202) 232-8656.

2. What campus unit(s) have the primary responsibility for these nonbibliographic machine-readable data bases? Check the statement most accurately describing your understanding of primary responsibility on your campus.

__5__ Library

__8__ Computer center

__3__ Shared between library and computer center

__7__ Other shared arrangement(s) involving library (briefly describe)

__14__ Other academic or research unit(s) (briefly describe)

__3__ Shared arrangements not involving the library. (briefly describe)

__1__ None.

3. Does the library use nonbibliographic machine-readable data bases available through online vendors?

__31__ Yes

__3__ No

IF YES, which vendors? Check as many as applicable.

__-__ ADP Network Services
__30__ Bibliographic Retrieval Services (BRS)
__-__ Data Resources (DRI)
__31__ Dialog Information Services
__9__ Dow Jones News/Retrieval
__-__ General Electric Information Services Mark III Service
__-__ Interactive Data Corp.
__17__ Mead Data Central (MDC) Lexis/Nexis
__3__ Sharp (I.P.) Associates
__16__ System Development Corp. (SDC)
__16__ Other(s) (indicate)

4. Does the library acquire nonbibliographic machine-readable data bases?

__12__ Yes

__22__ No

IF YES, in which format(s). Check as many as applicable.

__12__ Tapes
__-___ Disks
__1___ Cards
__-___ Other(s) (indicate)

5. In which of the following subjects areas does the library acquire nonbibliographic machine-readable data bases? Check as many as applicable.

__4___ Sciences (including agriculture and medicine)
__2___ Technology
__12__ Social Sciences
__3___ Humanities (including arts)

ENCLOSE LISTING(S) of data files acquired by the library.

6. Does the library have a collection development policy statement for nonbibliographic machine-readable data bases?

__4___ Yes

__30__ No

IF YES, enclose a copy

7. Does the library handle or store nonbibliographic machine-readable data bases?

__5___ Yes

__29__ No

If YES, enclose documents that describe handling procedures and storage facilities.

8. Does the library seek to make users aware of the availability of nonbibliographic machine-readable data bases?

__27__ Yes

__7___ No

If YES, by what means?

__5___ Bibliographic access through catalog of library holdings
__6___ Directory of resources available throughout campus. IF SO, briefly explain how library identifies existing machine readable files throughout campus.
__24__ Other (briefly describe)

9. Does the library offer services related to nonbibliographic machine-readable data bases?

32 Yes

2 No

If YES, check as many as applicable:

23 Referral outside the library
15 Advice on documentation
9 Advice on using programs
31 Library staff retrieval of data
1 Programming services by library staff
6 Library staff specifically assigned to assist users
5 Other services (briefly describe):

ENCLOSE descriptions of services such as user materials, staff training materials

10. What restrictions are applied to the selection and acquisition of nonbibliographic machine-readable data bases? Briefly describe, or enclose samples of guidelines and budgets.

-19 left blank
-14 made general statements
-1 reported, "committee"

FOR THE 1984 SPEC KIT ON THIS TOPIC, PLEASE BE SURE TO INCLUDE ANY DOCUMENTS YOUR LIBRARY CAN SHARE WITH OTHER ARL MEMBERS ON THE ADMINISTRATION OF MACHINE READABLE NONBIBLIOGRAPHIC DATA BASES. RETURN SURVEY AND DOCUMENTS BY MARCH 1, 1984, TO:

Nonbibliographic Data Bases Survey
SPEC Center/OMS
1527 New Hampshire Ave., N.W.
Washington, D.C. 20036

University of California at Berkeley

TEXTUAL AND NUMERIC DATA FILES

Primary Information Sources in Machine-Readable Form

A Report from the LAUC Ad Hoc Committee on Textual and Numeric Databases

Julie Kwan, Chair (UCLA)
Joyce Duncan Falk (UCI)
Sue Metzger (UCSD)
Margaret Mooney (UCR)
Sally Weimer (UCSB)

Librarians Association of the University of California

October 25, 1983

A PRELIMINARY THOUGHT

We are moving rather rapidly and quite inevitably toward a paperless society. Advances in computer science and in communications technology allow us to conceive of a global system in which reports of research and development activities are composed, published, disseminated, and used in a completely electronic mode.[1]

[1] Lancaster, F.W. Whither libraries? or wither libraries. *College and Research Libraries* 39:356, September 1978.

1. INTRODUCTION

The LAUC Ad Hoc Committee on Textual and Numeric Databases was constituted in December 1979 with representation from campuses in the north and south. The original charge to this committee was as follows:

1. to identify on each campus any office, department, or group offering access to any numeric and/or textual databases,

2. to collect policies relating to patron access to these services,

3. to recommend a course of action which could result in better access for our patrons without putting an unrealistic drain on library resources, and

4. to report at the Spring 1980 LAUC Assembly the progress made by the committee.

This committee met on 5 May 1980 with a draft report issued to committee members on 6 November 1981. The draft report commented on the difficulty if not impossibility of identifying all textual and numeric databases available at each campus. Written policy statements were extremely limited in number and in depth.

The draft report concluded: "Campus policies for dealing with numeric and textual data files need to be developed, and the libraries need to be an integrated part of these policies. If the library chooses not to become a data archive, the necessary links to the established campus archive need to be established and supported. Librarians need to be well informed on the availability of data files on and off campus, even if the library chooses not to be a collector of data tapes. Libraries need to collect written documentation and make it accessible to library patrons." The draft report did not include final recommendations.

In August 1982, the LAUC Executive Board, in discussing the problems of large ad hoc committees especially regarding limited travel budgets, reconstituted the committee with membership from the southern campuses. The new committee received a modified charge based on the interim action of the original committee. The new charge read as follows:

1. to recommend a course of action which could result in better access to and identification of non-bibliographic databases available without requiring extensive library resources, and

2. to report to the LAUC Assembly.

The new committee met on 21 January and 18 March 1983. The Spring 1983 Assembly was not held due to collective bargaining negotiations which precluded discussion of personnel issues, which would have constituted a significant percentage of the Assembly's agenda. Therefore, the new committee targeted its completed report for the Fall 1983 Assembly.

This report attempts to provide an overview of textual and numeric data files, their role in research and education, the role of the library in promoting access to them, measures for evaluating data files, factors affecting an institution's ability to provide access, and recommendations for action.

2. OVERVIEW OF TEXTUAL AND NUMERIC DATA FILES

This report is concerned with textual and numeric data files which are available in machine-readable form. Frequently, they contain primary rather than secondary data, and using the data may require manipulation by separate software packages. In addition, some secondary data sources and full text online journals fall into the scope of this report.

The proliferation of machine-readable data files over the last decade has presented libraries with the challenge of handling a new medium, information stored in machine-readable form. Since machine-readable data files are a relatively new concern for librarians, it is probably best to begin with a few definitions. Sue A. Dodd finds it useful to distinguish between primary and secondary data files.[2] Librarians are already familiar with secondary data files, for example, bibliographic databases produced in machine-readable form and accessible online by an online services vendor such as Bibliographic Retrieval Services (BRS) or DIALOG Information Services. While these secondary data files refer the user to the source of information, a primary data file contains the information itself.

Primary data files can be textual or numeric. One example of a textual data file is the Oil and Hazardous Materials-Technical Assistance Data System (OHMTADS) developed by the U.S. Environmental Protection Agency, which provides information on chemical substances to help in cleaning up hazardous spills.[3] The U.S. Bureau of the Census summary tape files are examples of numeric data files. Since numeric data files often provide raw, unanalyzed data, a separate database management package must be used to extract needed information from them. Two examples of software packages used to manipulate the data are SPSS (Statistical Package for the Social Sciences) and BASIS (Batelle Automated Search Information System).

The range of data files available is enormous, from the massive U.S. Bureau of the Census tapes to private files developed by individual researchers. The Summer-Fall 1982 issue of *Drexel Library Quarterly* (v. 18, nos. 3-4) provides a fairly comprehensive discussion of

---

[2] Dodd, Sue A. Toward integration of catalog records on social science machine-readable data files into existing bibliographic utilities: a commentary. *Library Trends* 30:336, Winter 1982.

[3] Heller, Stephen R. NIH/EPA Chemical Information System (CIS) physical and chemical databases. *Drexel Library Quarterly* 18:59, Summer-Fall 1982.

machine-readable data files available in both the scientific and social science disciplines.

The U.S. government is the largest single producer of numeric data files, with the Bureau of the Census, the Environmental Protection Agency, the Bureau of Labor Statistics, and many other agencies producing machine-readable data files.[4] In 1983, a committee was appointed within the federal government to consider providing machine-readable data files to depository libraries. This committee, the Ad Hoc Committee on Depository Library Access to Federal Automated Data Bases, has a target date of Spring 1984 for its final report. The final recommendations of this committee will undoubtedly be of interest to the academic community, and there may in the future be many changes regarding the availability of these data files to libraries which serve as government documents depositories.

Commercial vendors also produce and distribute data files, such as Standard & Poor's COMPUSTAT database of financial information about companies and R. L. Polk & Company's Profiles of Change database on housing. The distribution of data files is sometimes handled directly by the producers themselves. One large distributor of data files is the Inter-University Consortium for Political and Social Research (ICPSR), which was established in 1962. ICPSR is the major distributor of criminal justice data produced by the U.S. Department of Justice, as well as the major distributor of other social science data files to its member institutions.

Some files of textual or numeric data are available online, for example, the Registry of Toxic Effects of Chemical Substances, the Chemical Abstracts Service's Chemical Registry System (available through CAS Online), and the PTS files of economic forecasts and industry statistics. As many as 50% of all online databases are textual or numeric, according to John Fried and Gabor Kovacs.[5] These textual and numeric data files are frequently accessed by end-users, not necessarily through library or information intermediaries.

A growing trend is the emergence of the "electronic journal". Several journals are now available in this format, for example, a number of American Chemical Society journals are available full-text online through BRS. Most recently, Elsevier has mounted, also with BRS, recent volumes

---

[4] The enormous amount of information available in this format is evident through several directories, for example: *Catalog of public use tapes from the National Center for Health Statistics*. Hyattsville, Md., National Center for Health Statistics, 1980, DHHS (PHS)81-1213; *A directory of computerized data files*. Springfield, Va., Dept. of Commerce, National Technical Information Service, 1983; *Directory of federal statistical data files*. Washington, D.C., U.S. Dept. of Commerce, 1981; and *Directory of data files*. Washington, D.C., U.S. Dept. of Commerce, Bureau of the Census, 1979.

[5] Fried, John B. and Gabor J. Kovacs. Numeric databases in the 80s. *Drexel Library Quarterly* 18(3/4):2-10, Summer-Fall 1982.

of *Biochimica et Biophysica Acta*, *Brain Research* and *IRCS Medical Sciences*. The *Harvard Business Review* is available full-text on both BRS and DIALOG. The electronic journal, while still in its infancy, will play an important role in the future.

3. ROLE OF TEXTUAL AND NUMERIC DATA FILES IN EDUCATION AND RESEARCH

Textual and numeric data files are essential to the university's research and teaching mission. These files include sources of statistical and other numeric data, fact-answer oriented data, chemical and physical properties, and full-text reports of research. As the tools for recording and manipulating massive amounts of data have become readily available, scholars are able to investigate areas which previously had not been possible.

This technological revolution is affecting all disciplines. In the social sciences, for example, subject areas with available data files include: community and urban studies; conflict, violence, and wars; economic behavior and attitudes; elites and leadership; governmental structures, legislative bodies, and legal systems; political behavior; social behavior and institutions; organizational behavior; environment; health care; and the like. Similarly, issues of *Computers and the Humanities* and the programs of the International Conference on Databases in the Humanities and Social Sciences illustrate the development of data files in art, linguistics, music, and literature. In the physical sciences and engineering, quantitative data collection efforts are becoming extensive. In biology and medicine, files of genetic sequence data and genetic disorders are proving to be essential research tools, as are sources of health statistics, environmental hazards, and morbidity and mortality data.

Data files are also used increasingly for undergraduate and graduate coursework. Instructional packages on statistics and data analysis and instructional subsets of files on American politics, attitudes toward violence, the quality of American life, national elections, and other topics are available from distributors for classroom use in training students in research methodology.

The electronic journal promises an efficient means of disseminating results of research. It provides the full-text of reports and articles before or in lieu of the printed publication. This capability extends local collections of print materials and should ultimately offer increased availability of information. Members of the academic community should be allowed to take full advantage of information available in this form.

Currently, however, faculty and students are impeded in accessing data files. They face several dilemmas. First, it is difficult for them to know what sources are available. Records for data sets and services are not integrated into library records, the sources they normally consult for information. Because sources of data are scattered and because there is no uniform locator tool, users sometimes only become aware of them through

serendipitous methods. They will sometimes find information in the library, sometimes through organized units such as data archives or computing centers, but they may never know that the information they need is housed in a neighboring office or building. Or, the material may not be available locally at all. Second, there is no indexing available for the kinds of information available in individual data files. Users must refer to code books and descriptive manuals to determine if a data source includes the type of information needed. Third, faculty and research staff sometimes need specialized training or consultation to use many of these services, for example, with quantitative data analysis or chemical substructure searching. Often this training or consultation is difficult to obtain.

4. ROLE OF THE LIBRARY IN PROMOTING ACCESS TO DATA FILES

These textual and numeric data files play a central role in the mission of the academic library. According to the University of California *Academic Plan*:

> The library is the heart of a University. Knowledge and thought, the lifeblood of learning, accumulate in the library and circulate from it. Graduate education and research of the highest quality depend critically on what libraries can provide in access to primary sources and records.[6]

The library has a responsibility for acknowledging the information revolution represented by machine-readable data files and for assuring access to them. An awareness of these sources is important for the entire academic community, but currently this knowledge seems to be limited and dispersed. Even those sources which may be accessible through the library are often inadequately described at present. For example, do any of our cataloging or serial records indicate access to full-text materials online?

These sources often have qualities and characteristics different from traditional library materials and services and thus require new selection, access, and use policies. At present, libraries are not doing an adequate job in providing access to these sources, at least according to Blaise Cronin, who states:

> Librarians and information workers should wake up to the implications of the data bank phenomena or surrender the traditional reference function to service suppliers and information brokers.[7]

---

[6] University of California. Academic Planning and Program Review Board. *Academic Plan, 1974-1978*. Berkeley, California: University of California, 1974. p. 39.

[7] Cronin, Blaise. Databanks. *Aslib Proceedings* 33(6):243-50, June 1981.

There have been a number of recommendations in the literature encouraging an active role for libraries regarding data files. For example, Conger states:

> Libraries, as formalized interdisciplinary resource collecting units, whose function is to provide guidance to information in all its manifestations, are only now beginning to recognize their responsibilities in this area...The coordination of printed and computerized information resources is particularly valuable since work with one medium often impinges upon the work with the other.[8]

However, libraries do not have to work alone in this area. Within the University of California system, both the Los Angeles and Berkeley campuses have had social sciences data archives for several years. Still, much is needed, especially in other disciplines. Ferguson states:

> The natural division of specialized knowledge in libraries and computing centers suggests a collaborative effort. Such a collaborative effort should be able to provide the following elements: a data file and document budget, computing resources, data file tape management, statistical program packages, summary information about data files and programs, computing consultant services and library reference service.[9]

Perhaps at issue here is the evolving role of the library. Nina Matheson, in a recent report published in the *Journal of Medical Education*, proposes that health sciences libraries will play an integral part in a network which will integrate the enormous variety of data files within academic health sciences centers, including data files of library holdings, files of research and clinical data, and external bibliographic files.[10] While this report addresses the future of the health sciences library, certainly there is a parallel in the academic library regardless of discipline. As an example, academic libraries can and should play a greater role in promoting bibliographic accessibility to materials located in campus data archive centers, especially for those materials which might be important for a wide range of users, for which service is currently provided, and for materials which include unique and unduplicated information.

---

[8] Conger, Lucinda. "Data reference work with machine readable data files in the social sciences. *Journal of Academic Librarianship* 2:61, 1976.

[9] Ferguson, Douglas. Social science data files, the research library and the computing center. *Drexel Library Quarterly* 13:73,1977.

[10] Matheson, MW, Cooper JAD. Academic information in the academic health sciences center. Roles for the library in information management. *Journal of Medical Education* 57(10 pt. 2):1-93, October 1982.

5. MEASURES FOR EVALUATING TEXTUAL AND NUMERIC DATA FILES

The campus community must examine a variety of issues when assessing and evaluating the utility of textual and numeric data files, including consortium and contractual agreements, and for determining the specific roles of the library and academic community in providing access to these information sources. While some of these issues parallel standard collection development considerations, there are additional issues specific to this form of information resource which affect decisions of access.

1. *Subject Scope*: As with other materials, textual and numeric data files must be evaluated in relationship to the education and research interests of the academic community.

2. *User need*: The information needs of academic users must be examined to determine if primary data is needed in this format. For example, some users may need to manipulate primary data, rather than use already manipulated data. Conversely, some users may not need to deal with data manipulation and its attendant problems.

3. *Uniqueness of data*: As information sources proliferate, it becomes imperative to examine the uniqueness of data in this format. If the data is available in print, is it also necessary to have it available in machine-readable form? Why? There often are reasons to have access in both formats.

4. *Authority of data*: The origins of the data should be examined to assess authoritativeness and completeness. Data representing research in progress which might eventually undergo significant change is perhaps of limited value.

5. *Confidentiality of data*: Some textual and numeric data files may include proprietary or restricted information and consequently may be limited in terms of eligible user community. Some files may have suppressed information to insure confidentiality.

6. *Physical format*: In these times of rapid technological change, a variety of physical formats may be available. While primarily available on magnetic tape, some materials are beginning to be available in floppy disc or disc pack format.

7. *Online access*: Some sources, such as CAS Online, are available only online. For business and industry statistics, such as the PTS files, online access provides immediate forecasting capabilities. While most data sources are not available in this fashion, it is conceivable that more will become available as the need and demand for this type of information develops.

8. *Documentation, consulting and instructional support, hardware, and software*: Using textual and numeric data files requires adequate documentation so that users can extract needed information. Complex files may require staff to assist users with file manipulation. Hardware and software issues must also be addressed, including

availability of computing facilities and software within the campus community. This backup support may come from many parts of the campus environment.

6. FACTORS AFFECTING THE INSTITUTION'S ABILITY TO PROVIDE ACCESS

The following factors must be considered in determining the extent of the institution's ability to provide access to textual and numeric data files.

1. *Budget*: Access costs, whether for lease or purchase, may be considerable. Other cost issues include: initial processing, storing, and subsequent use of data files and costs for consortium fees, supporting materials, documentation, staff, space, and equipment. Some data files may be available for manipulation on an external computer (such as the NIH/EPA Chemical Information System) and may have attendant per-use connect charges. To what extent should these costs be supported from library budgets, campus departments, fees for services, etc.? Does whether or not these data files represent the only source of this information affect these decisions?

2. *Personnel requirements*: Consideration must be given to the staff time to provide the levels of service planned. Specialized subject knowledge beyond what is taught in the traditional library school curriculum is essential, just as managing a collection of information resources is not included in the computer science curriculum. Is the institution prepared to support this function?

3. *Space*: There must be sufficient storage space for these materials in the library and in other locations. In addition, there may be special environmental conditions required and security measures to protect against data loss and misuse.

4. *Organization and administration*: Organizational issues to be addressed include the question of who decides what files will be acquired, who will be responsible for bibliographic control, and what the division of labor among library departments and other campus units will be. The existence of other campus units (e.g., computing facilities, academic departments, organized research units) interested in these data sources and the degree of cooperation between the library and those units are primary factors in determining the library's role in providing or promoting access to these sources.

5. *Contractual agreements*: Contractual agreements may limit the number or category of users, thus restricting the institution's ability to provide service for all patrons. For example, a LEXIS contract may restrict usage to law school faculty and students; ICPSR files may be used only by affiliates of member institutions. Data files acquired by individual faculty members may not be accessible for other users without the faculty member's signature or consent.

6. *Communication*: The extent to which the library has adequate communication with the faculty, the campus contracts and grants offices, the campus public information offices, and research funding agencies will affect its ability to provide information about and access to these data files.

7. SUMMARY AND RECOMMENDATIONS

The rapid expansion of computer technology and the proliferation of textual and numeric data sources require a new examination of the library's role in providing services related to this medium of information. The LAUC Ad Hoc Committee on Textual and Numeric Databases recommends that University of California libraries, as dynamic institutions committed to the collection, organization, and dissemination of information generated in all forms, assume an active role in providing a range of services related to textual and numeric data files. Further, because the provision of these services involves considerable expenditure of both financial and human resources, the Committee recommends that a coordinated, University-wide approach be planned, whenever feasible, in order to insure maximum access to these materials.

RECOMMENDATION 1: THAT NEEDS ASSESSMENT STUDIES BE CONDUCTED FOR EACH CAMPUS.

The Committee recommends that campus libraries work closely with faculty, computing centers, data archives, and other appropriate campus departments to determine local needs for textual and numeric data files, both in the present and with projections for the future, and to establish campus policies and service plans based on these needs.

RECOMMENDATION 2: THAT CAMPUS ADVISORY COMMITTEES BE ESTABLISHED TO ADVISE ON LOCAL ACQUISITIONS AND SERVICES.

The Committee recommends that faculty advisory committees be established to provide input and recommendations regarding access to or purchase of data files. These campus advisory committees should be involved in the needs assessments mentioned above and in developing policy, acquiring budgetary support, addressing service-related issues, and serving as advocates for campus and university administrations. Evaluation factors mentioned in Section 5 above should be used in making these decisions with weighted consideration given to factors mentioned in Section 6.

RECOMMENDATION 3: THAT RESPONSIBILITY FOR TEXTUAL AND NUMERIC DATA FILES BE ASSIGNED TO SPECIFIC PERSONNEL.

The Committee recommends that each University Librarian assign authority and responsibility for coordinating library activities relating to machine-readable data files to specific personnel. Specific responsibilities should include:

1. coordinating the selection of data files or services and providing consulting and advisory services on use of those data files if these are provided through the library,

2. promoting awareness and use of textual and numeric data files,

3. keeping abreast of developments in this field,

4. serving as resource persons on matters related to textual and numeric data files, and

5. serving as liaisons between the library and other departments on campus.

RECOMMENDATION 4: THAT THE LIBRARY SERVE AS A CENTRAL REPOSITORY OF INFORMATION ABOUT TEXTUAL AND NUMERIC DATA FILES.

The Committee recommends that University of California libraries serve as central repositories of information related to textual and numeric data files. Specifically, campus libraries should collect:

1. published directories of data files and guides to the holdings of data archives which identify the existence and availability of these resources,

2. core collections of periodicals and monographs devoted to the utilization of these materials and to this developing field, and

3. code books, system manuals, and other documentation, if appropriate for campus needs.

RECOMMENDATION 5: THAT THE LIBRARY SERVE AS AN INFORMATION REFERRAL CENTER.

The Committee recommends that University of California libraries:

1. provide cataloging access to textual and numeric data files and related materials following the AACR2 and MARC format rules for machine-readable data files,

2. provide referral access for materials or services which do not fall within the above category, either through campus-wide directories of services or through linking records in local online systems, including descriptions of the eligible user populations,

3. plan staff development activities to promote awareness of textual and numeric data files, and

4. provide reference and referral services relating to textual and numeric data files.

RECOMMENDATION 6: THAT UNIVERSITY OF CALIFORNIA LIBRARIES COOPERATE ON A SYSTEMWIDE LEVEL REGARDING MACHINE-READABLE DATA FILES.

The Committee believes that systemwide coordination is essential for activities relating to textual and numeric data files. Library Council should insure that resources are appropriately shared throughout the campuses and develop a coordinated plan to provide the best services possible. The results of Recommendation 5 should be shared with all campuses by including cataloging records for data files in the systemwide online catalog, by distributing campus directories of data files and services, and by supporting communication of activities in this area.

RECOMMENDATION 7: THAT CONTRACTS BE NEGOTIATED WITH PROVISIONS FOR THE BROADEST POSSIBLE ELIGIBLE USER POPULATIONS.

The Committee recommends that campus and systemwide participation in contracts and consortia be developed whenever possible. As a minimum, the eligible user population should be defined at the campus level, rather than, for example, the department level. Cooperative sharing programs are recommended.

RECOMMENDATION 8: THAT FUNDS BE ALLOCATED TO PROVIDE ACCESS TO TEXTUAL AND NUMERIC DATA FILES ON EACH CAMPUS.

The Committee recommends that campus and systemwide administrations realize the value of textual and numeric data files for the research and instructional programs of the university and allocate funds to support appropriate acquisitions and services. Even with well developed policies and areas of responsibilities, the libraries, data archives, and computing centers need University-wide line item support for these services and materials.

RECOMMENDATION 9: THAT THESE RESOURCES BE INCLUDED IN LONG-RANGE PLANNING FOR THE UNIVERSITY OF CALIFORNIA LIBRARIES.

The Committee recommends that systemwide library administration incorporate appropriate textual and numeric data files in development plans. Previous development plans have excluded services such as online bibliographic services and audiovisual media; as a result, the development and support of these services has been hindered. Machine-readable data files should not also be ignored, whether they relate to magnetic tapes, online nonbibliographic search systems, or electronic publishing.

SUGGESTED READINGS

1 Borman, Lorraine and Richard Hay, Jr. The information transfer process on the university campus: The case for public use files. *Drexel Library Quarterly* 13:82-99, 1977.

2 Conger, Lucinda D. Data reference work with machine readable data files in the social sciences. *Journal of Academic Librarianship* 2:60-65, 1976.

3 Cronin, Blaise. Databanks. *ASLIB Proceedings* 33(6):243-50, June 1981.

4 Dennis, Jack. The relation of social science data archives to libraries and wider information networks. *Conference on Interlibrary Communications and Information Networks, Proceedings*, edited by Joseph Becker. Chicago: American Library Association, 1971. pp. 117-120. (Reprinted in *Reader in Machine-Readable Social Data*, edited by Howard D. White. Englewood, Colorado: Information Handling Services, 1977. pp. 165-174.

5 Dodd, Sue A. Cataloging machine-readable data files -- a first step? *Drexel Library Quarterly* 13:48-60, 1977.

6 Dodd, Sue A. Toward integration of catalog records on social science machine-readable data files into existing bibliographic utilities: A commentary. *Library Trends* 30:335-361, Winter 1982.

7 Ferguson, Douglas. Social science data files, the research library and the computing center. *Drexel Library Quarterly* 13:70-81, 1877.

8 Fried, John B. and Kovacs, Gabor J. Numeric databases in the 80s. *Drexel Library Quarterly* 18(3/4):2-10, Summer-Fall 1982.

9 Isaacson, Kathy. Machine-readable information in the library. *RQ* 22:164-170, Winter 1982.

10 Nasatir, Marilyn. The cataloging of machine-readable data files. Part II: A practical application of the developing principles. *Cataloging and Classification Quarterly* 1(4):41-62, 1982.

11 Robbin, Alice. The pre-acquisition process: A strategy for locating and acquiring machine-readable data. *Drexel Library Quarterly* 13:21-42, 1977.

12 White, Howard D. Libraries and access to social science data. in *Reader In Machine-Readable Social Data*. Edited by Howard D. White. Englewood, Colorado: Information Handling Services, 1977. pp. 175-194.

State University of New York at Albany

TO: Maxine Sitts
ARL/OMS SPEC Center

FROM: Bonita Bryant Bonita Bryant
Head, Collection Development

DATE: March 14, 1984

SUBJECT: Administration of nonbibliographic machine-readable databases in ARL university libraries

Attached is the SUNY at Albany response to your survey on this subject.

In addition to our responses to survey questions, you may be interested to know that the State University of New York at Albany University Libraries began to study the library's position as regards nonbibliographic databases with the appointment of a six-person taskforce in January of 1982. The taskforce's report was submitted to the Director of Libraries on February 15, 1983. The taskforce made a number of recommendations pointing toward library leadership on campus in identifying and providing access to non-bibliographic databases already acquired and those to be obtained in future. The taskforce report is enclosed.

Since the Fall of 1983 we have advertised unsuccessfully for a librarian who could serve half-time as Data Collections Coordinator and half-time as social sciences bibliographer. We have found that those qualified to perform the Coordinator tasks did not have the MLS required for the other segment of the position. We expect to advertise the position again, de-emphasizing need for experience with nonbibliographic databases, with the expectation that we will need to provide training for a neophyte who expresses interest and a creative approach to the task at hand. Copies of the first two positions advertised are attached.

Collection Development, 1400 Washington Avenue, Albany, New York 12222, 518/457-8555

State University of New York at Albany

VACANCY NOTICE

Position: Data Collection Coordinator/Sociology Bibliographer

Responsibilities: 1) Creates a Data Collection in the library of numeric and textual-numeric databases and documentation and coordinates its activities with those of the library's Computer Search Service, Campus Computing Center and campus research centers. 2) Develops collection policy for the Data Collection, with initial focus on the social sciences; evaluates, advises on, and recommends purchase of data files in support of reserach and instructional activities. 3) Provides reference service for the Data Collection; designs and develops data file access tools. 4) Serves as coordinator of data services functions and campus liaison to the Inter-University Consortium for Political and Social Reserach (CPSR). 5) Serves as Bibliographer for Sociology: selects, develops and evaluates subject collection; serves as liaison with faculty and students of this department and provides in-depth library instruction in this subject area. 6) Provides reference service including computer searching of sociology-related bibliographic and non-bibliographic databases. Reports to the Head, Collection Development. Library Faculty at SUNYA are expected to fulfill faculty obligations in the areas of contributions to the advancement of the profession and university service as well as specific library assignment.

Qualifications: MLS from an ALA accredited Library School. Graduate degree or equivalent in Sociology; strong background in statistics. Minimum of two years experience in academic libraries and/or data archive. At least one year's experience working with data files of numeric or textual-numeric databases and demonstrated competency therein. Ability to communicate effectively with students, faculty and library staff.

Salary & Rank: Commensurate with education and experience; salary from $19,000, depending on qualifications.

Terms & Benefits: Twelve months appointment; sick leave and annual leave @ 1.25 days each per month, fully paid major medical, hospitalization and dental insurance. Social Security coverage. TIAA/CREF or New York State Teachers Retirement available (employee contribution rate = 3%).

Campus & Community: The University at Albany, one of four university centers in SUNY, is located in New York's Capital city, 150 miles north of New York City, 150 miles west of Boston, 228 miles south of Montreal. The tri-city metropolitan area (Albany-Schenectady-Troy) has a population of approximately 300,000

Apply to: Send letter of application, current resume and names of three references to: Kevin McCarthy, Personnel Officer, University Libraries-Room 140, State University of New York at Albany, 1400 Washington Ave., Albany, New York 12222.

Inquiries should be received by October 10, 1983.

The State University of New York at Albany is an Equal Opportunity/Affirmative Action Employer. Applications from women, minorities and the handicapped are especially welcome.

THE UNIVERSITY AT ALBANY

VACANCY NOTICE

Position:	Social Science Bibliographer/Data Collection Coordinator
Responsibilities:	1) Shares bibliographer responsibilities for Social Sciences: selects, develops and evaluates subject collection; serves as liaison with faculty and students of this college and provides in-depth library instruction. 2) Provides reference service including computer searching of Social Science bibliographic and non-bibliographic databases. 3) Creates a Data Collection in the library of numeric and textual-numeric databases and documentation and coordinates its activities with those of the library's Computer Search Service, Campus Computing Center and campus research centers. 4) Develops collection policy for the Data Collection, with initial focus on the social sciences. Evaluates, advises on, and recommends purchase of data files in support of research and instructional activities. 5) Provides reference service for the Data Collection; designs and develops data file access tools. 6) Serves as coordinator of data services functions and campus liaison to the Inter-University Consortium for Political and Social Research (CPSR). Reports to the Head, Collection Development. Library Faculty at SUNYA are expected to fulfill faculty obligations in the areas of contributions to the advancement of the profession and university service as well as specific library assignment.
Qualifications:	MLS from an ALA accredited Library School. Graduate degree or equivalent in Social Sciences. Experience in academic libraries and/or data archive; experience working with data files of numeric or textual-numeric databases preferred. Ability to communicate effectively with students, faculty and library staff.
Salary & Rank:	Commensurate with education and experience; salary from $17,500 depending on qualifications.
Terms & Benefits:	Twelve months appointment; sick leave and annual leave @ 1.25 days each per month, fully paid major medical, hospitalization and dental insurance. Social Security coverage. TIAA/CREF or New York State Teachers Retirement available (employee contribution rate = 3%).
Campus & Community:	The University at Albany, one of four university centers in SUNY, is located in New York's capital city, 150 miles north of New York City, 150 miles west of Boston, 228 miles south of Montreal. The tri-city metropolitan area (Albany-Schenectady-Troy) has a population of approximately 300,000.
Apply to:	Send letter of application, current resume and names of three references to: Library Personnel Officer, University Libraries - Rm. 140, State University of New York at Albany, 1400 Washington Avenue, Albany, New York 12222. Inquiries should be received by January 31, 1984.

The State University of New York at Albany is an Equal Opportunity/Affirmative Action Employer. Applications from women, minorities and the handicapped are especially welcome.

UNIVERSITY LIBRARIES • 518/457-8353
STATE UNIVERSITY OF NEW YORK AT ALBANY • 1400 WASHINGTON AVENUE • ALBANY, NEW YORK 12222

State University of New York at Albany

University Libraries

Taskforce on Non-bibliographic Databases

FINAL REPORT

Submitted to the Director of the Libraries: February 15, 1983

Taskforce Members:

Meredith Butler (liaison)
Roger Gifford
Richard Irving
Sally Knapp
Amy Vegter
Ann Wood (chair)

TABLE OF CONTENTS

1.	Introduction	p.	1
2.	Task Force Objectives	p.	1
3.	Definition of Non-bibliographic Databases	p.	2
4.	Literature Review	p.	3
5.	Current Use of Non-bibliographic Databases at SUNYA	p.	9
6.	Present Services for Non-bibliographic Databases in the Libraries and on Campus	p.	10
7.	Future Need for Non-bibliographic Database Services at SUNYA	p.	11
8.	Identifying Alternatives for Meeting the Need for Access to Non-bibliographic Databases	p.	13
9.	The Role of the Libraries--Future	p.	18
10.	Task Force Recommendations	p.	19
	Bibliography	p.	27
	Appendix A: The charge to the Task Force on Non-bibliographic Databases	p.	29
	Appendix B: Summaries of the Task Force Meetings with Richard Alba and Barbara Wolfe	p.	31

1. Introduction

The Task Force on Non-Bibliographic data bases was appointed on January 12, 1982 and was charged with exploring the role the University Libraries should play in accessing, acquiring housing and/or servicing non-bibliographic databases (see Appendix A for the complete charge to the Task Force). Since that time the group has met 14 times, conducted an extensive review of the literature, and interviewed two persons from the campus (Richard Alba, Director of the Center for Social & Demographic Analysis, May 6, 1982 and Barbara Wolfe, Director of the Computing Center and Associate Vice President, April 8, 1982: See appendix B.)

The members of the Task Force are:

Meredith Butler (liaison)
Roger Gifford
Richard Irving
Sally Knapp
Amy Vegter
Ann Wood (chair)

2. Task Force Objectives

The Task Force defined its objectives as follows:

a. Define non-bibliographic databases.

b. Conduct a review of the literature on non-bibliographic databases.

c. Identify current use of non-bibliographic databases on campus.

d. Identify services related to non-bibliographic databases presently provided by campus units (e.g. libraries computer center, Center for Demographic Research, departments).

e. Projected needs for non-bibliographic databases on campus in the 1980's.

f. Identify alternatives for meeting those needs.

g. Identify the possible roles the Libraries might take in meeting these needs.

h. Recommend future steps for the Libraries and campus to take.

i. Identify the impact of the recommendations on budget, staffing, and space.

3. Definition of Non-Bibliographic Databases

Databases are collections of data that can be accessed and/or manipulated by computer. Bibliographic databases are a type of reference database which provide access to literature through bibliographic citations or abstracts.

Non-bibliographic databases comprise the spectrum of all other databases to which a library user might wish to have access. They may contain either source or referral data. The Smithsonian Science Information Exchange is an example of a non-bibliographic referral database which contains descriptions of ongoing government-sponsored research projects.

Source databases can also be classified according to whether their data is primarily textual (e.g. Encyclopedia Brittanica), textual-numeric (e.g. Predicast databases containing summaries of statistical forecasts and time series), or numeric (e.g. U.S. Census tapes, or scientific survey data). The Structure and Nomenclature Search System (SANSS) is an example of yet another type, a "properties" database, in which chemical compounds are retrieved by "drawing" their structure.

As the above examples illustrate, non-bibliographic databases may contain either source or referral data. They may be available either online or offline, and may or may not be amenable to statistical manipulation. Some are accessed only through sophisticated online searching systems, but a greater number are accessed offline using user-written software or general-

purpose statistical packages such as SPSS (Statistical Package for the Social Sciences).

4. Literature Review

The Task Force members undertook a rather extensive literature review to investigate the issues and problems surrounding the topic of non-bibliographic database services in academic libraries. Some of the key findings are summarized below.

The definition of non-bibliographic databases adopted by the Task Force is broad in scope. It includes referral, numeric, textual-numeric, property, and full text databases. However, this literature review mainly addresses the need for services concerning numeric databases, for the following reasons:

1. The overwhelming proportion of the information which the Task Force received, from both literature and interviews, dealt with numeric databases.
2. A survey of on-line non-bibliographic databases found that 256 online non-bibliographic databases surveyed 192 were numeric (Wanger and Landau, 1980, 172).
3. The proliferation of numeric databases has had a great impact on the information needs of students and researchers,especially in the social sciences. At the same time, there has not been a corresponding development of services and facilities, in either libraries or other institutions on campuses, which provide access to computer manipulable numeric data.
4. From a librarian's point of view, numeric databases pose some unique reference problems regarding access to and manipulation of the data.

The solutions to the problems of servicing machine-readable (m-r) data files in academic settings have included the development of separate data archives: data archives located in the university library, a special library, a research institute, or computer center. Some m-r data archives are nothing more than data repositories while others provide a whole range of services including reference assistance in locating appropriate data and programming assistance in analyzing it (Byrum and Rowe, 1972, 339). Administratively data archives are attached to libraries (Yale) (Salter, 1975, 12), part of research institutes (Michigan) (ICPSR, 1982/83, 1), the computer center (Princeton) (Rowe and Ryan, 1974, 7-15) or set up as a special library (Wisconsin-Madison) (Robbin, 1982, 407). Some institutions have divided the functions of the archive creatively between two or more of the above units (University of British Columbia) (Ruus, 1982, 397).

There has been some controversy in the literature as to whether collections of m-r data should be housed in archives or special libraries. Those who favor archives stress the fact that m-r survey data have in common with the materials of a more traditional archive the status of primary or raw evidence (Miller, 1976, 409). Those who favor the designation of special library (Robbins, 1982, 407) tend to emphasize the reference functions provided at their institution. However, certain functions are unique to m-r data such as editing or "cleaning" the data and constant editing and updating of codebooks (Dennis, 1970, 9).

Major archives now collecting and disseminating social science data include the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan (Traugott, 1981, 1066+), the Roper Center attached to the University of Connecticut and Williams College (Delucia, 1978, 535+), the International Data Library at Berkeley, and the Institute for Behavioral Research at York University. Data files are

also available from numerous govermental agencies such as the U.S. Bureau of the Census and the Social Security Administration. Local campus service archives normally hold data from one or more of these sources as well as some locally produced data sets (Byrum and Rowe, 1972, 339).

The data archive movement, according to White, (1977, 175-194) developed in the 1960's as a means to save a unique type of primary source data and make it available for secondary analysis. Whether academic libraries should add m-r numerical data to their holdings and engage in the corresponding archival and reference functions is still a matter of debate and controversy. Hayes and Becker (1970, 710-714) in their discussion of a library-based center feature the pros and cons of such an arrangement in a special list, showing that strong arguments can be made on both sides.

White summarizes the problems associated with a library-data archive merger. He thinks, however, that libraries have a definite role to play to make information about data more widely available. Access to m-r files is of two kinds--physical and intellectual. Intellectual access in the form of codebooks and other tools of access should be made readily available and this is economically feasible for libraries. He also argues that libraries are in a unique position to tie codebooks of source data and monographs based on that data together through their card catalogs. In addition, libraries provide systematic and objective reference services and access to information that does not depend on knowing the researcher who has the source data stored in his files or office. A university library formalizes and depersonalizes access to information and the user sees the library as the information access center on campus. In contrast, physical access to the source data is not so readily available because of the high costs involved.

Rowe and Ryan (1974, 7-15) also see a role for libraries in the collection and servicing (i.e. copying data, extracting data, editing

data, etc.) of m-r data. Although they focus on census tapes, their argument can be extended to cover the general topic of m-r data. They agree that users are trained to view the university library as a center for information and that libraries have a commitment to public service, continuity of collections and/operations, and longer hours of service than most other campus facilities. They identify major issues associated with provision of m-r services and describe services provided at UCLA and Princeton as examples of two contrasting roles a library might take in providing these services.

One message from these and other articles in a recent issue of Library Trends devoted to "Data Libraries for the Social Sciences" (v. 30 no. 3, Winter 1982) "is that there is no single administrative structure for services related to social science data files nor for numeric information systems that is ideally suited for all institutional settings" (Adams, 1982, 377).

Even though there is no agreement on exactly who should provide and where m-r data services should be provided, there is general agreement on the types of services which should be provided at the campus level. An ideal level of service is described by Dennis (1970, 9-10) to include, in addition, to the usual library functions of acquisitions, cataloging, indexing, referencing, storing, maintaining, reproducing, and distributing information; an m-r data service should also edit or clean data, update and edit codebooks, provide computer programs for use in file handling and statistical analysis, give consultations on the proper use of the programs, and furnish general education to student and faculty users.

Various of the sources consulted describe lesser levels of service than indicated above which could be offered. Ruus (1980, 29-33) article is especially helpful in its discussion of user services for a data library based on Rothestein's three levels of reference service: minimum, middling, and maxium. A conservative or basic service data library would be one which

purchased m-r files on request only; maintains data files as they are received without much or any checking, cleaning, etc.; maintains codebooks as they are received from the supplier; and offers some basic consultation services in data processing and statistical procedures. Difficult problems are referred to experts in the computing center, statisticians, etc.

Middle and liberal levels of service are described in some detail and depend both on fiscal resources and longevity of the data library. The author concludes that at the outset, a data archive/library should concentrate on user services in order to cultivate an active and supportive user community.

Included in effective services are the provision of reference tools which provide access to the data files. Kolp (1980, 33-38) discusses the three types of tools which appear to be more useful to users of m-r data files:

1. data catalogs describing the contents of individual data archives and data libraries.
2. directories describing the contents of more than one archive, or directories within special topical areas.
3. periodicals which attempt to report information on the holdings of social science data archives.

The author then goes on to describe features of each type of publication in some detail. At the end of his article he cautions, however, that the print medium may now be out of date as a means of communication in this field. Online systems at individual archives and projects whereby archives exchange data descriptor tapes certainly hold the promise of a much improved data reference system.

In terms of online services which might be provided as a part of reference or information retrieval services in a library, Mignon (1980, 181-187) identifies four types of requests which are reasonable for a trained librarian to handle:

1. convenience search - - factual, quick, might be faster than searching a printed document.

2. aggregate search - - collection of facts, variety of variables to be collected and used together, can be arranged in a single output.

3. reorganized data search - - information required mathematical operations on the figures, but arithmetic rather than statistical procedures.

4. cross tabulation tables - - looking at variable relationships using simple statistical manipulations.

To provide this kind of service and to give reasonable reference services for offline m-r data, Mignon also identifies several skills a librarian should have:

1. knowledge of the characteristics of numeric databases, the information they contain, how they order the information, and the uses typically made of them.

2. recognition of the information needs which are most appropriately satisfied by numeric databases.

3. knowledge of how to find out if there is a numeric database that will satisfy a specific need and where to direct the user in order to obtain access to that file.

4. recognition of the kind of numeric database needs that cannot be satisfied in a general reference environment, and the ability to suggest the services of an appropriate data archivist or subject specialist.

The techinical services aspects of documentation and bibliographic control of m-r data files are discussed by Byrum and Rowe (1972, 339-346). They present a practical plan for the documentation and control of data sets in m-r format, one which does not depend on the computer skills of the data librarian or on the computer systems available to the librarian. They believe their scheme provides sufficient flexibility to allow for the variety of ways in which data libraries are organized and administered, recognizing that a library may have full charge of the data as well as its accompanying documentation or it may only hold the catalog records while the data is handled either by a computer center or by some independent facility. The four elements of documentation and control are: a standard catalog entry, a data abstract of data description form, content documentation or codebook, and records of

the physical and logical characteristics of the data set. The plan can be implemented, according to the authors, without extensive financing.

A history of the efforts at providing standards for bibliographic control of m-r data is provided by Dodd. She notes that chapter nine "Machine readable Data Files (MRDF)" is in AACR2 and that a MARC format has been developed for m-r data files. As of 1982 "no bibliographic utility had incorporated the MRDF/MARC format into its system, but two local applications have been realized--at the Social Science Data Library, University of North Carolina and at the Inter-university Consortium for Political and Social Research." (Dodd, 1982, 352). She provides examples of a MARC-formated record, a catalog entry, and a data-abstract for an m-r data file.

5. Current Use of Non-bibliographic Databases at SUNYA

One information need that the Taskforce identified is that of a survey of current campus use of non-bibliographic databases. Without a survey it is difficult to estimate the current level of use. However, we do have some less precise indications; requests for numeric data tapes and searches, and for relevant software training in the library and computer center respectively. In addition, the Center for Social and Demographic Analysis has begun an inventory of the large social science data sets on campus. This inventory indicates some need for information and services in this area.

Almost all graduate departments of the College of Social and Behavioral Sciences and the professional schools list courses which stress statistical methodology and computer application.

6. Present Services for Non-bibliographic Databases in the Libraries and on Campus

At SUNYA, there is no one designated organization which collects and disseminates information about non-bibliographic databases. Currently, the scope of the Libraries' involvement in non-bibliographic databases is limited. There is an ICPSR bibliographer who coordinates the University membership in the Inter-University Consortium for Political and Social Research and at the present houses the codebooks and catalogs associated with the latter institution in her office. The Libraries also possess miscellaneous catalogs and codebooks from other institutions and of other data sets. The collection of catalogs and codebooks is not uniformly cataloged or stored in one given location. In addition, SUNYA's libraries subscribe to several newsletters dealing with the topic such as SS Data.

At present, on-line involvement in non-bibliographic databases, coordinated through the Information Retrieval (IR) Unit, consists of searching databases on demand. These searches are reimbursed by the patron to cover some or all of the costs involved. The IR Unit is also gradually integrating statistical and directory databases available through existing subscribed-to vendors into its services. We anticipate that current Unit policies will have to be revised to accommodate increased demand for non-bibliographic database services.

Other services for non-bibliographic databases on campus outside the Libraries include the technical consulting services offered by the Computing Center to faculty/staff and graduate students working on individual research projects. These services cover the use of the UNIVAC operating system, programming languages, statistical packages, tape conversions, personal computing, typesetting, and data communication problems. Undergraduate student instruction is handled by course instructors and teaching assistants.

Many departments and schools have centers such as the Humanities and Fine Arts computing laboratory where there are a number of mini or micro computers and a terminal or terminal's connected to the University's mainframe UNIVAC computer.

The Center for Social and Demographic Analysis in the Sociology Department is the local campus depository for New York State 1980 Census Data on tape. Services include providing tabulations of the data tailored to individual specifications. The Center also has the NORC General Social Survey 1982 and the related codebook available for anyone on campus to use. There are possibly other units which would turn up in a more detailed inquiry.

With the exception of the Center, the services provided are those of physical access and help with computing problems not with the organization and provision of the tools for access to non-bibliographic databases. The Center, provides access to two very clearly defined data sets and could, with the investment of money, be expanded to include more sets; however, it does not yet control the universe of university owned non-bibliographic databases already on campus or have organized access to information about other possible data sets.

7. Future Need for Non-bibliographic Database Services at SUNYA

Regardless of the current level of known use for access to numeric databases at SUNYA there are numerous indications of increasing demand for access to such data in the future. First, a revolution in computer technology has already reduced the spatial, temporal, and financial restrictions which previously inhibited the collection, organization, dissemination, and analysis of quantitative data. The result has been an exponential increase in the amount of numeric data produced, and in research based on numeric data. The increase in numeric data and in knowledge based on that data has been

accompanied by the need, and demand, for intellectual and physical access to the data. If previously collected data is easily accessible, secondary analysis of that data is possible and saves the researcher the time and cost of collecting and organizing his/her own data.

Second, as the costs of computers decrease more researchers and academic departments will be purchasing their own equipment and will place additional service demands on those campus units which provide access to databases, assistance in programming, and instruction in the use of software program packages.

Third, reductions in research funding availability will encourage researchers to rely on secondary analysis of existing data and will make imperative adequate means of access to that data.

Fourth, the use of data and computers for classroom instruction is considered a necessary component of the social science learning process, along with training in methodology and data handling. Richard Alba, Director of the Center for Social and Demographic Analysis, estimates that even now 75 to 80% of the graduate students on campus become computer literate during the course of their studies. In addition, students in secondary and elementary education are being exposed to computer technology to such an extent that it is likely the next generation of university students will be computer literate before they arrive on campus. Thus automated systems are likely to be a preferred mode of access to information including numeric databases.

Finally, some data, particularly that issued by the Federal government, will increasingly be issued only in machine-readable form.

8. Identifying Alternatives for Meeting the Need for Access to Non-bibliographic Databases

The trends detailed in section seven indicate that the University will need to include services for non-bibliographic databases in campus automation planning. If this need is to be met, a coordinated University response which addresses the following questions should be developed:

1. What steps can the university take to provide adequate physical access to non-bibliographic databases?
2. Is there a need for a central storage and service center on campus for offline non-bibliographic databases?
3. Which campus unit(s) should be responsible for acquiring, organizing and giving access to the data sets?
4. What kinds of data should be collected and retained?
5. What steps have to be taken to provide adequate intellectual access (bibliographic control) to non-bibliographic databases?
6. Where will information regarding software programs be located? Who will provide assistance in the use of software programs and what will the level of assistance be?
7. What personnel requirements will arise from the provision of these services?

Possible responses to each of these questions are addressed below:

1. Physical access: Some of the needs for physical access to non-bibliographic databases are already being met through the campus membership in the Inter-University Consortium for Political and Social Research (ICPSR) and the provision of access to some online databases by the Libraries through Dialog and BRS. Leaving aside the question of the necessary financial resources, expansion of these services *per se* is fairly easy. Memberships in additional archives (e.g. the Roper Center) could be purchased and contracts could be signed with other commercial online database vendors.

2. Storage and service center: Our interviews and readings have

indicated that there is a need to establish a campus data collection of offline data sets. Factors pointing to this need are outlined above and reiterated here in a slightly different form.

(a) Increasingly, data, especially from government sources, are being made available only in machine-readable form. Researchers will need to have local access to some of these data sets.

(b) Acquiring data costs money, and in a period of fiscal constraints, building a centralized collection of frequently used data sets makes the most efficient use of financial resources.

(c) A depository is needed for locally generated data, especially since funding sources frequently stipulate that data collected for a research project become the property of the institution rather than the individual researcher. Barbara Wolfe, Associate Vice President for Computing Services, reported to the Task Force that the Computing Center is sometimes left with locally generated data that no one claims. An often suggested alternative to a local data collection: that of encouraging researchers to contribute their data to appropriate national data archives should be advocated for major data sets that attract a national audience, but this alternative is inappropriate for data which is locally oriented, or which is being used locally whether generated on campus or acquired from outside. Sets of data which are used for instructional purposes also may need to be stored locally. A local archive would collect and organize these types of data and would complement national archive collections by receiving data needed for campus use from many sources and ensuring that it is available to local users.

3. Responsibility: Our research has found no one best answer to the question of which unit should be responsible for acquiring, organizing and giving access to data sets. The configuration on a campus depends

on local factors. Responsibility may be assigned to the computer center, the library, a research institute, a separate data library, or be divided between two or more of these units. The SUNYA campus unit, or units working together, which can best provide acquisition, deposit and access services for locally held data needs to be designated and supported with sufficient financial and personnel resources.

4. Collection policy: Two approaches to collection development for non-bibliographic databases are often mentioned: the preservation of all raw data which is thus preserved for secondary analysis; and the retention of only those data sets, the use of which justifies their upkeep. Policies for selection, retention and discarding non-bibliographic data sets need to be developed based on the teaching and research needs of the University and the need to preserve some locally generated data sets.

5. Intellectual access: A recurring theme in the literature is the inadequate state of bibliographic control (cataloging) of machine-readable data both within and between institutions holding such collections. AACR II rules for the descriptive cataloging of machine-readable data sets and the MARC record format will help to alleviate this problem of access. However, other access tools to individual data sets are necessary. These include:

(a) data abstracts or data description forms, which could be kept in loose-leaf binders or which might eventually exist in machine-readable form, (b) content documentation codebooks, (c) references to published research based on the data, (d) records of the physical and logical characteristics of the data set (Byrum and Rowe, 1972, 344-345).

Some data archives provide catalogs of their holdings (e.g. ICPSR's Guide to Resources and Services). In addition, there are directories which describe the contents of more than one archive, newsletters, agency bulletins and other publications which provide information regarding the existence and availability of data sets. It makes sense to gather, organize, and house all of these materials in one central well publicized location on campus.

Considerable progress has been made in developing major national social science data archives. The organization and functions of the local level archive and its links into an emerging national network are, however, not as yet well defined. Certainly a campus information center containing the information necessary to access and use locally held data sets and the tools to locate and access data sets held by other archives and distributors would serve as a campus focal point in the development of a national data archive network.

6. Provision of software program services: Increases in the use of numeric data will require the provision of services regarding software programs. Specific services which will be required include: "comprehensive documentation, user orientation, introductory courses on commonly used statistical packages and program languages, and extensive programming consultation." (Ruus, 1980, 32) On this campus the Computing Center provides instruction in the use of statistical packages and programming languages, and programming consultation. However, one of the conclusions of the SUNYA Committee on Directions for Academic Computing (June 1980) was that "There is a major need for social science consultation by personnel familiar with social science data and data analytic methods." (p.7) Instructional programs regarding statistical packages and programming languages will need to be expanded as the

university community acquires computers which can handle additional social science software packages such as OSIRIS IV. More complicated program services e.g. "writing a formula for the retrieval of data from a complex database; or creating a program to provide uniform access to and manipulation of aggregate data across several censuses" may need to be developed (Ruus, 1980, 32).

The campus unit which can most appropriately provide these services needs to be designated and supported both financially and with personnel.

7. Personnel Requirements: Personnel requirements in the Libraries will be dependent upon the type and level of service decided upon.

The appropriate role for a library data center would likely follow Ruus' suggestion that it include "in-depth searching of the literature, compiling bibliographies, and doing a fair amount of the user's work for him." Staff would be available for some consultation but more complicated questions regarding manipulation of the data would be referred to experts in other departments (Ruus, 1980, 30).

To provide this level of service the manager of a library data information center should have at least basic computational skills and statistical training.

A final determination regarding how services associated with non-bibliographic databases are divided among campus departments is a question which should be resolved at the university level. What is clear is that there will be a need for expert personnel on campus designated to provide technical and substantive advice regarding the manipulation of machine-readable data. If the provision of those services are left unorganized then large groups of users will be excluded from their use. If the services are organized at a university level the effect will be not only to make the service available to more users but to stimulate research that would otherwise not be undertaken.

9. The Role of the Libraries--Future

Our interviews and readings have indicated that libraries are being asked, especially by social scientists, to take responsibility for organizing, cataloging and providing access to offline data sets. In separate interviews with the Task Force, Richard Alba of the Center for Social and Demographic Analysis and Barbara Wolfe of the Computing Center both indicated that they felt that the University Libraries at SUNYA should assume these functions. They expressed five reasons:

1. Organizing and cataloging are things libraries know how to do.
2. The library is a central institution on campus and people are trained to look to the library to satisfy information needs.
3. Even though the Computing Center is also a centralized campus facility, the staff tend to see themselves as helping people process or manipulate data rather than helping to locate or develop data.
4. Libraries have experience in developing acquisition policies and building collections.
5. The future of institute, such as the Institute for Social and Demographic Analysis, is unstable and depends on students and scarce university funds for its day to day operations.

An additional argument in favor of library involvement in m-r data services not mentioned during the interviews is that libraries contain other research materials which augment and compliment m-r information sources. As also noted in our review of the Literature whether academic libraries should add machine=readable numeric data to their holdings is still a matter of debate and controversy. It would appear to depend on the local configuration of institutions on each campus where such data are best acquired. It is, however, apparent that our informants and other authors agree that whether the SUNYA Libraries set up a data archive or not they should become the campus information center for locally held data sets, the holdings of other archives, and for online services. All agree that with minimum additional

resources for purchasing documentation, space, and technical services, this service is economically feasible.

Beyond serving as an information center for social science numeric data, the Taskforce agrees with our local informants that the goal of the Libraries should be to develop a local data archive in the Libraries. It would serve the purpose of minimizing the duplication of both services and data on campus. Neither a specialized institute or the Computing Center is as well equipped to service such a campus need as are the Libraries. It cannot be done, however, without widespread debate and agreement on a campus level.

10. Task Force Recommendations

The Taskforce recommendations which follow take into account these considerations. They are presented as: Short range recommendations addressing what the University Libraries can do immediately to meet campus information needs. Long range recommendations which we hope will be the subject of campus-wide discussion, debate, and decision on the most appropriate way to organize the services needed for university-owned non-bibliographic databases.

Short range:

1. The Libraries should set aside space for a browsing and circulating collection of numeric social science database documentation.
   a. The space requested by the Taskforce from the Space Planning Committee is Room 310 and the adjoining two offices.
   b. Transferred to this space from other parts of the Main Library should be the ICPSR codebooks and other materials now stored in the ICPSR bibliographer's office, data archive catalogs and union lists of database holdings, newsletters and bulletins of major

archives, and other codebooks from the Reference Collection, the Stacks, and the Periodical Room. The probable number of such transfers affecting CASM is minimal (2 or 3 periodical titles and not more than 50 monographs).

c. Assign a location symbol to the Room and to the cataloged portion of the documentation collection.

d. The immediate equipment requirements for Room 310 and the offices include, in addition to regular office equipment for one of the offices:

-3 five-shelf bookshelves for ICPSR materials to be located in the second office
- book stacks for the cataloged documentation for Room 310
- tables and chairs for using the collection

2. The Libraries should begin systematically to collect information about, and documentation for, numeric, textual-numeric, full-text and referral databases which support the teaching and research needs of the University community. These materials will be added to the collection. A budget and fund code should be established to acquire these materials.

3. A person on the Libraries' staff should be designated coordinator of the collection. The coordinator should:

a. acquire database documentation and provide reference service for it. Initially reference service might consist of keeping designated office hours.

b. coordinate campus ICPSR membership.

c. act as liaison to relevant faculty, schools, intitutes, and Computing Center.

d. administer the funds designated for the collection.

e. coordinate the activities of the unit with other units of the Libraries, especially the Information Retrieval Unit.

f. train and coordinate other faculty of the Libraries who have assigned responsibilities in the unit.

g. develop ideas for a collection development policy to be written subsequent to campus-wide approval of the development of data archive.

h. disseminate information about services offered by the unit.

4. Online numeric, referral, full-text and textual-numeric databases available through vendors with which the Libraries have contracts should be considered for access by the Information Retrieval Unit in consultation with the Reference Services Division and the coordinator for numeric databases. IR policies may have to be adjusted to allow for this.

5. Additional contracts for online database searching should be considered where existing contracts are not adequate to meet the instructional and research needs of the university community.

Long range:

Recommendations for long range activities will be discussed in five categories: (1) general, (2) collection policy, (3) storage/organization, (4) service/reference and (5) required resources.

1. General: Raise with Vice-President for Reasearch and Educational Development and other appropriate persons or groups on campus the issues noted in this report in order to obtain campus-wide agreement on the future servicing and support for non-bibliographic databases. Specifically, the need for the development of a campus data archive, it's proper location, sources of funding, and personnel need to be considered. What follows is the Taskforce's proposal to develop a Data Archive.

2. Collection Policy: The Libraries should be designated the campus unit responsible for collecting and managing publicly available numeric, referral, full-text and textual-numeric databases.

   a. The Libraries should routinely acquire and make available documentation for non-bibliographic databases. A collection development policy for documentation acquisition should be developed.

   b. Membership in producer archives (e.g. the Roper Center, ICPSR) should continue to be coordinated through the Library.

   c. The Data Archive, in close cooperation with academic units involved, should collect, and develop a collection policy for the acquisition of, offline numeric data sets. At first the policy should be limited to acquiring data sets on demand. The collection should also be limited to locally developed and/or locally used data.

   d. The Data Archive should develop policies for the ownership and retention of data sets acquired with state funds, grant funds and/or by gift or deposit.

   e. The Information Retrieval Unit should develop policies to select online services and should monitor new developments and the use of existing services. Factors to be considered in evaluating the services include the availability of personnel on the Libraries' staff to learn new systems; and funds for training them and to acquire the necessary system documentation.

   f. The Libraries should not attempt to acquire or access simulation, modeling or other online databases which are designed for use by end-users.

g. Documentation on software (e.g. SPSS Manuals) designed to manipulate offline numeric databases should be collected by both the Data Archive and units elsewhere on campus which provide software consultation and programming assistance.

3. Storage/Organization:

a. Documentation for non-bibliographic databases should be stored and organized by the Unit which will provide service for it. For example, documentation for BRS and Dialog databases accessed by the Information Retrieval Unit should be maintained by that unit. Documentation for non-bibliographic databases not accessed there should be maintained in Data Archive collection.

b. Documentation on offline data files should be fully cataloged and kept together in the Data Archive. This would involve assigning a location symbol preceeding the call number (similar to the 'DOC' collection). Most documentation has at least member copy in OCLC.

c. The responsibility for physically storing locally held data files should be the Campus Computing Center's.

d. Archival and backup copies of all tapes should be made. Procedures to ensure the perpetual integrity of locally held data sets should be worked out to maintain both the archival copy and a working copy (e.g. copy the tapes at designated intervals).

e. Access to the local data archive by dataset name, tape number and file number should be maintained in the Libraries.

Option a: Create a separate card or online catalog of locally held data sets. Using this option, the catalog will have to be updated each time tapes are copied.

Option b: Interfile cards for locally held data sets in the

main card catalog (AACR II, Chapter 9) while maintaining a separate list which links tape numbers to the catalog information.

4. Service/Reference: Reference service should be provided for non-bibliographic databases.

   a. All reference librarians should be alerted to whatever level of service the Libraries implement, and should be prepared to direct patrons to the proper unit (Information Retrieval Unit for online databases, Data Archive for the documentation collections, card or online catalog of locally held databases, or the Computing Center for complicated software assistance).

   b. Liaison with the Computing Center and the departments and research institutes most concerned should be actively maintained by the coordinator so that persons inquiring in any of these places can be directed to the proper location.

   c. The Data Archive coordinator, in conjunction with other relevant officials, should explore the possibility of obtaining grants or other external funding to implement the archive.

   d. Data Archive and Information Retrieval personnel should provide specialized reference service for non-bibliographic databases. Services from online databases to which the Libraries have access should be limited to retrieval of data and simple functions such as aggregating, sorting and cross tabulation of data. Additional online statistical manipulation will be provided only if a person can be trained or added to the Libraries' staff. Statistical manipulation of data sets acquired for persons (but not available to the Libraries online) will be the responsibility of the patron. Limited software consultation, location of data, and referral to other departments and centers offering in depth services should also be offered.

e. Policies on services and charging for online databases (SUNYA Personnel only or the community at large?) need to be examined. The IR policies already developed for online bibliographic databases may not be adequate to handle online non-bibliographic databases.

f. Policies on the use, loan and redistribution of offline data sets by or to other institutions or non-SUNYA individuals need to be developed. (ICPSR, for example, does not allow its tapes to be redistributed (copied or loaned) to persons or intitutions outside member institutions. Locally developed databases, however, with the permission of their depositor, may well be used by, or loaned or sold to, other institutions or individuals).

5. Required resources: The resources needed to develop an archive can be divided into four basic categories: a) space, b) personnel and training, c) equipment, and e) funds.

a. Space: The immediate space requirements of the archive include room to store documentation, space in the Computing Center and/or a State vault for tapes and space for a terminal or terminals connected to the campus mainframe computer. All but Computing Center space can be satisfied by the assignment of Room 310 and the two adjacent offices to the Archive. Other space needs may be identified in the future.

b. Personnel: It is essential for the data archive coordinator as well as library staff who provide in-depth reference service to have training in statistics and familiarity with the use of software packages. The Libraries should hire such a person or persons, or identify persons already on the staff who have, or can obtain, the appropriate training.

Some clerical support will be needed for typing, filing and the physical maintenance of the documentation collection.

Technical Services personnel will have to be trained to catalog both the documentation and the datafiles acquired by the archive. The cataloging of documentation poses no special problems. However, datafiles are an unfamiliar media to our staff and a considerable amount of special knowledge will be required in order to catalog data sets to the level of service that we envision.

c. Equipment:

- A terminal and printing facilities.
- Other equipment to be identified in later stages of planning.

d. Funds will be needed for:

- memberships in producer archives, online service contracts, purchase of data sets and documentation, and state vault rental fees.
- personnel: the coordinator, part-time clerical assistance, and reference staff.
- terminal(s), printer(s)
- tapes

Some of this money may be recoverable from grants or from the schools and departments for whom data and documentation is acquired.

BIBLIOGRAPHY

Adams, Margaret O'Neill
1982 "Models of data library development and information systems services: an overview." Library Trends. 30:3:377-381 (Winter)

Byrum, John D. Jr. and Judith S. Rowe
1972 "An Integrated user-oriented system for the documentation and control of machine-readable data files." Library Resources & Technical Services. 16:3:338-346 (Summer).

"Data libraries for the social sciences."
1982 Library Trends. 30:3. (Winter). Special Issue.

DeLuca, Donald
1978 "Archival data resources and analysis services of the Roper Center." Journal of Conflict Resolution. 3:535-544 (September)

Dennis, Jack
1970 "The Relation of social science data archives to libraries and wider information networks." Washington, D.C., Office of Education. 17p. (ERIC ED 057-856)

Dodd, Sue A.
1982 "Toward integration of catalog records on social science machine-readable data files into existing bibliographic utilities: a commentary." Library Trends. 30:3:335-361. (Winter)

Hayes, Robert M. and Joseph Becker
1970 Handbook of data processing for Libraries. New York, Becker and Hayes.

1974 Handbook of data processing for Libraries. Los Angeles, Melville Publishing Co. 2nd Ed.

Inter-university Consortium for Political and Social Research.
1982/83 Guide to resources and services. Ann Arbor, ICPSR, annual.

Kolp, John G.
1980 "Reference tools for machine-readable data files." IASSIST Newsletter 4:2:33-38. (Fall).

"Machine-readable social science data."
1977 Drexel Library Quarterly. 13:1 (January). Special issue.

Mignon, Edmond
1980 "Numeric databases in the professional libranianship curriculum: implications for behavioral and social sciences librarians." Behavioral and Social Sciences Librarian. 1:3, 181-187 (Spring)

Miller, Warren E.
1976 "The Less obvious functions of archiving survey research data." American Behavioral Scientist. 19:4:409-418 (March-April).

Reader in machine-readable social data.
1977 Ed. Howard D. White. Englewood, Colo., Information Handling Services.

Robbin, Alice
1982 "The Data and program library service: a case study in organizing special libraries for computer-readable statistical data." Library Trends. 30:3:407-434. (Winter).

Rowe, Judith S. and Mary Ryan
1974 "Library services from numerical databases: the 1970 census as a paradigm." College and Research Libraries. 35:1:7-15. (January).

1977 "Ibid." Reprinted in Reader in machine-readable social data. Edited by Howard D. White. Englewood, Colo., Information Handling Services. 195-205.

Ruus, Laine G.M.
1980 "User services in a data library." IASSIST Newsletter. 4:2:29-33 (Fall)

1982 "The University of British Columbia Data Library: an overview." Library Trends. 30:2:397-406. (Winter).

Salter, Billie I.
1975 "The Social Science Data Archive of Yale University" EDUCOM Bulletin. 10:3:12-15. (Fall).

Traugott, Michael W.
1981 "The Availability of resources for public policy analysis from the ICPSR." Policy Studies Journal. 9:7:1066-1069. (Summer).

Wagner, Judith and Ruth M. Landau
1980 "Nonbibliographic on-line data base services." Journal of the American Society for Information Science. 31:3:171-180. (May)

White, Howard D.
1977 "Libraries and access to social science data." In Reader in Machine-readable social data. Edited by Howard D. White, Englewood, Colo., Information Handling Services. 175-194

State University of New York at Albany

TO: Roger Gifford
Richard Irving
Sally Knapp
Amy Vegter
Ann Wood

FROM: Joseph Z. Nitecki, Director of Libraries

DATE: January 12, 1982

SUBJECT: Task Force on Non-Bibliographic Data Bases

I appreciate your willingness to serve on this Task Force which Ann Wood has agreed to chair.

A major role of the University Libraries at SUNYA is to support the teaching and research programs of the University. Teaching and research in all academic fields on this campus, as elsewhere in the country, is increasingly involving both the use of computers and the use of statistical data. This use of computers has resulted in the building of files of data in the form of both bibliographic and non-bibliographic databases. In academic libraries, bibliographic databases, such as those provided by Dialog and BRS, have been in common use for the last five years as tools providing researchers appropriate citations to published and unpublished sources. Non-bibliographic databases (presently defined as those providing directory, dictionary and/or statistical information online or those involving data not directly available through an online service) have not had as much direct impact on libraries as yet. They are, however, transforming the way research is done in the sciences and social sciences and to a lesser extent in the humanities. Universities are responding to this transformation by providing computing facilities and services, often scattered throughout the campus, to allow both the development of non-bibliographic databases as integral parts of research projects and also provide for the storage and reuse of the data. Outside of libraries, database archives, such as the Interuniversity Consortium for Political and Social Research and the Roper Center, have been developed to store, service and redistribute databases for use in secondary analysis. In addition, bibliographic database vendors are beginning to offer access to statistical and other non-bibliographic databases and government organizations are producing huge files of machine readable statistics (e.g., U.S. Census) and distributing them in various ways to secondary users, many located at academic institutions.

University Libraries, UL 123, 1400 Washington Avenue, Albany, New York 12222, 518/457-8541 or -8540

This Task Force should address itself to the question of the role the University Libraries should play in accessing, acquiring, housing and/or servicing non-bibliographic databases. Issues to be considered include:

1. What kinds of files or services for, or access to, non-bibliographic databases already exist or are planned for implementation in the near future on this campus? (An inventory of this type of information is already in process through the Office of the Vice President for Research.)

2. In light of the above information about the present situation on campus, should the University Libraries consider non-bibliographic databases as another form of information which we should access or acquire, document, and service as we do other sources of knowledge? If so, how should we define non-bibliographic databases? Or should we attempt only to serve as a referral point to other units (e.g., the Computer Center, the Institute for Social and Demographic Analysis, etc.) on campus which provide, or will provide, the requisite services? Is there some point of involvement between the two extremes at which the University Libraries can provide useful services?

3. Assuming some involvement is decided on, how much and what will the impact be on budget, staffing (cataloging, reference, etc.) and space, at various levels of involvement? Should we consider developing another unit within the Libraries, similar to the Information Retrieval Unit, which would have microcomputers and terminals connected to the SUNYA computer, etc., and be staffed by persons trained to aid patrons, get access to, and use non-bibliographic databases? What are the advantages and/or disadvantages of establishing a separate unit? Where would such a unit fit in the Libraries organizational structure?

4. Is there any potential for the development and implementation of joint programs, related to non-bibliographic databases, with the Computer Center, the Office of the Vice President for Research, or any campus units and applying for grants to implement them?

By April 15, 1982 I would appreciate your report addressing the above questions of the Libraries' role vis-a-vis non-bibliographic databases and making your recommendations as to the type and level of involvement the University Libraries should have in accessing, acquiring, housing and/or servicing these databases. Included in your report should be a brief reading list for colleagues, acquainting them with the issues you have been addressing and with which they may be dealing, depending upon our level of involvement with non-bibliographic databases.

Thank you.

JZN:lr

Appendix B: Summaries of the Taskforce's Meetings with Barbara Wolfe and Richard Alba.

April 8, 1982
Barbara Wolfe, Director of the
Computing Center and Associate Vice-President

Present: A. Wood (Chair), M. Butler, R. Gifford, R. Irving
S. Knapp - Visitors: J.Z. Nitecki, B. Wolfe

After introductions and a short description of the Task Force's charge, the group had a fruitful discussion with Barbara Wolfe on the possible roles of the Library and the Computer Center in dealing with non-bibliographic databases. Some of the points which emerged from the discussion are noted below:

1. To whom does dissertation data in m-r form belong? Usually to the writer of the dissertation. Copyright laws define this.

2. The Task Force is familiar with Judith Rowe at Princeton and with IASIST.

3. AACRII does have rules to catalog machine readable data. According to B. Wolfe this function (that of organizing and providing access to m-r data) should be handled by the library.

4. The proper physical location for the data itself is the Computer Center. However, the refreshing of tapes and making of backups should be done by the library or other agency (it is a clerical job) via remote terminal.

5. A basic question raised by B.Wolfe which has to be answered before we should begin to collect and preserve m-r data is -- Is the data of sufficient value to the institution to invest in it and it's maintainence? There must be standards and/or a collection policy.

6. An extension of our present reference service would be to service government m-r data. More and more material is going to come out from the Government in m-r as opposed to printed form. B. Wolfe feels that a Computer Center is not the place to service this material. A library is already set up to collect and provide access to such data. Computer center people are into computing - not collecting and organizing data.

7. B. Wolfe will support the Library if we decide to build up an m-r collection and to provide access to it. The Computer Center is equipped to give technical assistance on software packages and on how to manipulate data.

8. It was agreed that the next step should be for the Chair of this task force to contact the Chair or entire membership of the Committee on Research Computing (William Lanford, Physics, Chair) of the Presidential Commission on Computing to acquaint them with our concerns; assertain faculty perceptions of possible need for a library role in handling m-r data and the need for an archive for locally generated data.

May 6, 1982
Richard Alba, Director of the
Center for Social and Demographic Analysis

Present: A. Wood (chair), M. Butler, R. Gifford, D. Irving, S. Knapp
A. Vegter - Guest: Richard Alba

Prof. Alba, after a brief introduction to the charge of the taskforce, explained his view of the ideal relationship between the library and the computer center. Within two decades the two organizations should overlap. The print and non-print breakdown will no longer be valid. Computer literacy will force the Library to expand its horizons. Already 75-80% of graduate students are computer literate. Libraries should begin now to find ways to satisy needs which are just beginning to be expressed.

The library should have an important role in archiving data. Libraries are especially important because they know how to catalog. In the Center for Social and Demographic Analysis, Alba is collecting information about social science data sets on campus, but he and his aids don't know how to analyze and catalog the sets for retrieval. The library can centralize information on social science data on campus. A room or rooms should be provided with information about data sets, documentation, on them, and computer hardware (terminals) to access them.

The computer center has people trained in computing - and the staff there view themselves an information processers. The cc staff doesn't have substantive knowledge of any academic field, but they feel their job is to assist people to manipulate data from any field.

In the future less and less information will be printed - and more and more will be available only in machine - readable form. It therefore is necessary for the library to begin to deal with machine readable records.

It is extremely likely that decentralization of computer facilities on campus will take place in the next 10 years: 1) A single mainframe will not be comprehensive enought to satisfy all needs; 2) computing is becoming much cheaper.

To legitimatize the Library's role in archiving and cataloging information about machine readable data, the issue should be raised at the university level. Alba would like to see a reorganization and a partial merger between the Computer Center and the Library. His center is unstable - depending as it does on student personnel and scarce University resources. The Library and Computer Center are more stable institutions - and with legitimacy at the University, level resources could be obtained to organize to handle machine - readable data.

Current demand for machine readable data should not be taken as a measure of future demand. An entirely new generation of computer - literate students will arrive on campus within a decade.

The Library's immediate goals in addressing these future needs should be to provide space (a room), cataloging, and documentation for machine readable files. Alba's Center has begun an inventory to identify large data sets on campus. As far as developing an archive is concerned, criteria to judge what is worth keeping should be developed. For the Social sciences, one can begin by 1) identifying major polls and 2) often cited surveys.

The nature of research will be different in 10 years. Secondary analysis of data collected by others will become much more important because there is not as much money available for primary data collection. Also most faculty now can't do original programming so they are restricted somewhat in the complexity of analysis and number of databases they can use at one time. This will change with the new generation of students. The lack of money for primary data collection will also incline people to pool resources for centralized data file collection building. Alba also believes that computer services, just as library services have been in the past, should be provided free. There should be no charges.

Alba sees software package provision and maintenance as a function of the computer center.

AW:ea
2/11/83

Resource Fact Sheet

Date: 10/79
Number: FS0015

SOCIAL SCIENCE DATA ARCHIVE

The Social Science Data Archive is the repository and reference center at Yale for machine-readable social science data. Operating within the Social Science Library, the Archive extends the special services offered by the library into the area of computer-readable data files stored on magnetic tape. Such files include census data, data from public opinion polls and surveys, studies in economic behavior, and legislative roll call data, among others.

The Archive is the successor to a data collection formed by the Political Science Research Library in 1959. In 1972, when the Political Science Research Library was incorporated into the Social Science Library, the Archive was established to maintain responsibility for the data collection and to provide data services.

RESOURCES AND SERVICES AVAILABLE THROUGH THE SSDA

The Archive currently maintains over 500 data holdings on magnetic tape. Among these are:

* Contemporary and historical census enumerations, including:
    * Public Use Samples and Summary Tapes from the 1970 Censuses of Population and Housing
    * Public Use Samples from the 1960 Censuses of Population and Housing
    * Current Population Surveys, 1968-date
    * 1972 Census of Governments
    * Historical census materials, 1790-date
* Surveys of economic behavior and attitudes, such as:
    * National Longitudinal Surveys of Labor Market Experience (Herbert Parnes and the Center for Human Resource Research at Ohio State U.)
    * Panel Study of Income Dynamics (James Morgan and the Survey Research Center at U. of Michigan)
    * Consumer Expenditure Survey, 1972-1973, Diary Survey and Interview Survey Detailed Public Use Tapes (Bureau of Labor Statistics)
* Data on economic processes, including:
    * International Financial Statistics (International Monetary Fund)
    * CITIBASE, formerly NBER Economic Time Series Databank (Citicorp)
* Education studies, including:
    * National Longitudinal Study of the High School Class of 1972 (Reseach Triangle Institute)
    * Carnegie Commission National Survey of Higher Education: Faculty Study (Everett Ladd, Seymour Lipset, Martin Trow and the Social Science Data Center at UCONN)
* Events data, such as:
    * World Event/Interaction Survey (WEIS) (Charles McClelland)
* Data from legislative and deliberative bodies, including:

    * Roll call voting records for the U.S. Congress, United Nations, and foreign nations

* Election and primary studies, such as:

    * American National Election Studies and Panels, 1948-1978, (Center for Political Studies, Survey Research Center, U. of Michigan)

    * Numerous election studies from other nations

* Election returns, including:

    * State Level Congressional, Gubernatorial and Senatorial Election Data and State-Level Presidential Election Data for the United States, 1824-1972 (W. Dean Burnham, Jerome M. Clubb and William Flanigan)

    * Numerous other United States historical election returns, 1788-date

* Electoral data, including:

    * Congressional District Data Books

    * County and City Data Books

* Public opinion polls on political matters, such as:

    * Euro-barmeter 3, May 1975, and Euro-barometer 6, Oct/Nov 1976 (Jacques-Rene Rabier and Ronald Inglehart)

    * European Community Studies, 1970, 1971, 1973 (Jacques-Rene Rabier and Ronald Inglehart)

* Studies of social institutions and behavior, such as:

    * National Survey of Familiy Growth, Cycle I (National Center for Health Statistics)

    * School District Data Tape (Combined 1970 Fourth Count School District Data Tape and Elementary-Secondary General Information Survey) (Applied Urbanetics, Inc. for National Center for Educational Statistics)

* Community and urban studies.

* Data on conflict, agression, violence and war

* Elites and leadership data

* Studies of the environment and natural resources

* Data on governmental structures, policies, and functions

* Social indicators data

The Archive provides services related to machine-readable social science data. These services include the acquisition of data files which have been requested or which update files already held by the Archive, reference services to help users locate data both within and outside the Archive's holdings, and advice in using documentation and programs with Archive data.

The Archive administers the Yale membership in ICPSR (Inter-University Consortium for Political and Social Research). Data may be obtained from or deposited with ICPSR by contacting the Data Reference Librarian. Also, the OSIRIS III program package is made available at YCC through the Archive's ISPSR membership.

HOW TO ACCESS THESE RESOURCES

Social Science Data Archive holdings are stored at the Yale Computer Center and are available to all faculty members, students, and staff at Yale University. Indi-

viduals from other academic and non-profit institutions should apply to the Archive's Data Reference Librarian. All data reference materials such as codebooks, tape descriptions, directories, etc. are available at the Social Science Library (SSL) at 140 Prospect Street. Most of the Archive's data holdings are cataloged like other library materials, and cards describing the cataloged machine-readable files are interfiled in the library card catalog. Also, the Archive maintains an indexed Guide to Resources and Services describing all of the data files in the Archive collection. Further assistance may be obtained from the Archive staff:

JoAnn Dionne, Data Reference Librarian, 436-3412 (SSL)
Sandra Aivano, Staff Resource Specialist, 432-4080 (YCC)

OTHER PROGRAM AND DATA RESOURCES

Although the Yale Computer Center (YCC) offers users a wide variety of program packages, programming languages and other facilities, YCC cannot support all existing program and data resources. Several organizations outside of YCC, support programs and/or data which these organizations make available to YCC users. In general these resources are highly valuable, but are of interest to a more narrow scope of users than could justify YCC support. Currently the organizations providing such resources to YCC users are the Social Science Data Archive, the Social Science Computing Facility, the Roper Center, and Edunet.

SOCIAL SCIENCE DATA ARCHIVE (SSDA)

The Social Science Data Archive is the repository and reference center at Yale for machine-readable social science data. Operating within the Social Science Library, the Archive extends the special services offered by the library into the area of computer-readable data files stored on magnetic tape. Such files include census data, data from public opinion polls and surveys, studies in economic behavior, and legislative roll call data, among others. Refer to Resource Fact Sheet #FS0015 entitled SOCIAL SCIENCE DATA ARCHIVE.

THE SOCIAL SCIENCE COMPUTING FACILITY (SSCF)

The Social Science Computing Facility, or Statlab, was established in July 1977 to consolidate and facilitate statistical computation for the social science community at Yale. In order to improve computation facilities for the general social science community, the SSCF has collected and cataloged existing computer software, modified (and in some cases developed from scratch) general purpose statistical packages, and advertised their availability. The Statlab has developed extensive documentation of program characteristics for its program library and sponsors seminars to explain the intricacies of this growing library. Refer to Resource Fact Sheet #FS0014 entitled SOCIAL SCIENCE COMPUTING FACILITY.

THE ROPER CENTER

The Roper Center for Teaching and Research, now on the Yale campus, provides a valuable research resource for the social sciences. The Center houses the oldest and largest archive of sample survey data in the world.

Established in 1946 as the Roper Public Opinion Research Center, the archive contains raw data and supporting documentation from thousands of individual studies carried out nationally and abroad. More than 100 domestic and international survey organizations have contributed their data to the public domain through the cooperative efforts of the world's major research organizations and the Roper Center in recent years. Many data sets are from nations other than the United States, including Great Britian, France, Germany, Denmark, Canada and Japan as well as other countries in Europe, Asia, and Latin America. Refer to Resource Fact Sheet #FS0013 entitled THE ROPER CENTER.

EDUNET

EDUNET is a national network of colleges and universities that makes possible the sharing of computer-based resources. It provides the marketplace and organizational framework that permit colleges, universities, and other research and teaching institutions to exchange many desirable and often unique resources.

University of British Columbia

Selected Sections

PROCEDURE

Data Library Procedures Manual

written by

Laine Ruus

Computing Centre

UNIVERSITY OF BRITISH COLUMBIA
6356 Agricultural Road
Vancouver, B.C., Canada V6T 1W5

April 1982

in kit

Table of Contents

I. PURPOSE 1

II. MANDATE 2

III. ADMINISTRATION 3
- A. RELATIONS WITH THE LIBRARY AND COMPUTING CENTRE 3
  - 1. Computing Centre 3
  - 2. Library 3
  - 3. Other Departments And Faculties 4
- B. PERSONNEL 4
  - 1. General Policy Statement 5
    - a. The Librarian 5
    - b. The Programmer 6
    - c. The Library Assistant II 6
  - 2. Job Descriptions 7
  - 3. Training 8
- C. BUDGET 9
  - 1. Data Library Acquisitions Account 10
    - a. Getting 10
    - b. Spending 10
    - c. Kick-backs 11
    - d. Processing Invoices 11
  - 2. Annual Reports, Etc. 12
    - a. Collecting Monthly Mount Statistics 12
    - b. Printing A Year-end Mount Report 12
    - c. Others 13
  - 3. Supplies, Equipment, Stationery, Etc. 13

IV. A SHORT CHAPTER ON THE LONG SUBJECT OF CONTRACTS 15
- 4. Contracts For New Data Files Being Deposited 16
- 5. Contracts For Data Files To Be Disseminated Elsewhere 17
- 6. Moral Contracts With Users 18

V. COLLECTIONS, IDS, ETC. 19
- A. INTRODUCTION TO FILE MAINTENANCE 19
- B. INTRODUCTION TO THE COLLECTIONS 20
  - 1. CHUM:CATALOG 20
  - 2. DLIB:CATALOG 21
  - 3. CHUM:ARCHIVE 21

VI. THE CARE AND FEEDING OF THE TAPE COLLECTION 22
- 4. Acquiring New Tapes 22
  - a. To Purchase A New Tape Or Tapes 22
  - b. Enter It Into CHUM:CATALOG 23
  - c. Enter It Into TAPESHARE 23
  - d. Assigning Rack Number, And General Housekeeping 24
  - e. Assigning Tape Labels And Tapenames To Tapes ..25
  - f. Tapes Purchased From Elsewhere 27
  - g. Tape Cleaning 27

h. Tape Rewriting 28
i. ARCHIVEal Tapes 28

VII. THE CARE AND FEEDING OF THE MRDF COLLECTION 30
A. ACQUISITIONS POLICY 30
B. ACQUISITION OF MRDF: 31
1. The Identification And Location Of Mrdf. 31
C. ACQUISITIONS PROCESS 33
D. BLANKET ORDERS 34

VIII. WHEN A NEW DATA SET ARRIVES 35
1. Accompanying Documentation 35
2. What Kind Of Tape? 35
3. Edit CHUM:CATALOG 36
4. Check The Tape Documentation For The Size Of The Files 36
5. Edit Dlib:catalog 37
6. RUN THE MGR PROGRAM 38
7. BACKUP 39
8. Mount Test 39

IX. DOCUMENTATION 40
A. TECHNICAL DOCUMENTATION 40
B. THE CARE AND FEEDING OF THE CODEBOOK COLLECTION 41
1. Types Of Codebooks 41
2. Duplicate Copies Of Codebooks 42
C. SECONDARY DOCUMENTATION 43
1. SPIRES:DATALIB 44
2. Creating New SPIRES Records 44
3. Field Descriptions 45
4. Subject Classification 52
5. Cutter Numbers 55
6. Procedure Examples 55
a. Short Update 55
b. Long Update 56
c. Finding An Unknown Key 56
d. Processing The File 57
7. Field Associations 57
D. TERTIARY DOCUMENTATION 58
1. Technical Processing Of Codebooks 59
a. Codebooks 59
b. Blue Cards 59
E. DISTRIBUTION FILE 60
F. WHERE TO FILE EVERYTHING 60
G. REMOTE DOCUMENTATION 60
H. SUPPLEMENTARY DOCUMENTATION 61
I. NEWSLETTER PRODUCTION FROM SPIRES DATABASE 61
J. PRODUCTION OF THE COMFICHE DATA LIBRARY CATALOGUE ...62
c. Introduction 63
d. All Catalog Entries 63
e. Subject Index 63
f. Index 63
K. ANNOUNCEMENTS TO S S DATA 63

X. DATA DISSEMINATION .....69.....65

XI. THE CARE AND FEEDING OF THE REFERENCE COLLECTION 67
A. ACQUISITION AND PROCESSING OF REFERENCE TOOLS 67
1. Serial Ordering 67
2. Cataloguing Serials 67
3. Serial Check-in 68
4. Serial Claiming 68
5. Binding Serials 68
6. Ordering Monographs 68
7. Cataloguing Monographs 69
8. Binding Of Monographs 69
9. Transfers 70
10. Sliding Locations 70
11. Withdrawl Of Library Materials 70
12. Superseded Microfiche 70
B. THE CARE AND FEEDING OF THE SPIRES:BIBLIO FILE 70
1. Input 71
2. Creating Input Records 71
3. Adding Records To The Database 72
4. Adding Many Records To The Database 72
5. Updating Records In The Database 73
6. Finding An Unknown Record Number 73
7. Processing Records 74
C. THE CARE AND FEEDING OF THE SPIRES:ORDERS FILE 74

XII. DATA FILE CLEANING 76
A. POLICY 76
B. MULTI-PUNCHED DATA FILES 76
1. MEG - Card Image Checking 77
a. Purpose 77
b. Availability 77
c. Type Of Routine 77
d. How To Use 77
e. Method 78
f. Restrictions 78
2. To Write Multipunched Data Into A File 78
3. MULT - List Multipunched Columns 79
a. Purpose 79
b. How To Use 79
4. PEEK And PURGE - Column Binary Cleaning 79
a. Purpose 79
b. Availability 79
c. Type Of Routine 79
d. How To Use 80
e. Method 80
f. Timing 81
g. References 81
5. SPRAY - Spraying Multipunched Columns 81
a. Purpose 81
b. Availability 81
c. How To Use 81
d. Messages 83
e. Restrictions 83
f. Subprograms Required 84
g. Sample Program 84
C. PROCEDURE FOR CREATING OSIRIS IV DATA FILES 84

1. Input Format 85
2. Wild Code Checking 85
3. Recoding The Data 86
4. Creating Codebook Records 87
5. Codebook Marginals 88
6. Final Codebook 89
7. Records Management 89
D. CREATING SPSS SYSTEM FILES 90
E. UN-SPSSING SPSS SYSTEM FILES 90

XIII. USER SERVICES 93
A. POLICY 93
B. CIRCULATION 93
1. Circulated Materials 93
2. Borrowers 94
3. Date Due 94
4. Overdues 94
5. Returns 95
C. DECENTRALIZED CODEBOOK COLLECTIONS 95
D. DLIB:DOC 95
1. Maintenance Of DLIB:DOC 96
E. CLASSROOM USE 96
F. PUBLIC RELATIONS 96
G. INTRODUCTORY LECTURE 96

XIV. EXTRA-INSTITUTIONAL RELATIONS: ASSOCIATIONS, ETC. 98
A. POLICY 98
B. CANADIAN 98
C. IASSIST 98
D. IFDO 98
E. WAPOR, ETC. 98

XV. DLIB:BATCH 99
DLIB:BATCH - Enter And Save A Batch Run 99

XVI. MGR PROGRAM 100
A. PURPOSE 100
B. INTRODUCTION 100
C. CONTROL STATEMENT SUMMARY 100
1. Basic Commands 100
2. Special Reformating Commands 101
3. Modifying The Store And Save Commands 102
4. Global Commands 102
D. CONTROL STATEMENT SYNONYMS 103
E. ENTERING BATCH RUNS AT THE TERMINAL 103
F. CONTROL STATEMENT EXAMPLES 104
1. Storing Files From Labeled Tapes 104
a. One Input Tape 104
b. Two Or More Tapes At Once 104
2. Storing Files From Unlabeled Tapes 104
a. One Input Tape 104
b. Two Or More Tapes At Once 105
3. Storing Files From Disc 105
4. Overwriting Files 105
5. Backing Up 105

6. Fetching Files From Data Library Tapes 106
7. Documenting Files On Data Library Tapes 106
a. Specific Files 106
b. Large Numbers Of Files 106
8. Large Runs 106
9. Documenting A Tape Already Mounted 106
G. EFFICIENT USE 107

XVII. PROGRAMMING SUPPORT FOR DATA PROCESSING 108
A. PURPOSE 108
B. BASIC TAPE STORAGE PROCEDURES 108
C. CANSIM 108
1. CANSIM Cumulative Base 108
2. CANSIM University Base 109
D. CITIBASE 109
E. CRSP 109
1. Stock Files 110
2. Bond Files 110
F. COMPUSTAT 110
G. IMF 110
H. SPSS FORMAT FILES 111
I. IMAGES 111
J. MAINTAINING DATALIB AND BIBLIO UNDER SPIRES 111

XVIII. PROGRAMMING SUPPORT FOR PUBLIC PROGRAMS 114
A. DLIB:IMF 114
B. DLIB:SDB 114
C. DLIB:CRSP 114
D. DLIB:CRDI, DLIB:CRMI, DLIB:CRDR, DLIB:CRMR 114
E. DLIB:CPST 115
F. DLIB:GRP 115
1. General 115
2. Compilation 115
G. OBSOLETE PROGRAMS AND FILES 117
H. OTHER FILES UNDER ID DLIB 118

XIX. PROGRAMMING SUPPORT FOR GRP MAP FILE 119
A. PURPOSE 119
B. DISCLAIMER 119
GRPCLN - Removing Null Lines 121
GRPBILD - Building A Segment File 122
GRPINDEX - Building A Higher Level Index 123
GRPWALK - Re-ordering Indexes 125

INDEX 126
T=6.684

$SIG

II. MANDATE

The Data Library's MANDATE, as outlined in the original proposal which created the Data Library in 1972 (see Kennedy & Stuart-Stubbs) is to support research and teaching activities of the academic community of the University of British Columbia.

Both parent institutions of the Data Library, namely the Computing Centre and the Library, provide their services to the whole academic community of the University, with no restrictions as to either the academic discipline nor academic status of the individual user. I have always assumed, and it has never been disputed, that this service mandate, although not stated in so many words, has applied to the Data Library as well. It is however, not possible to be all things to all people - there are limits to the services we will provide, as there are limits to the amount of service provided by either the Computing Centre or the Library. Service policy will be described elsewhere.

III. ADMINISTRATION

A. RELATIONS WITH THE LIBRARY AND COMPUTING CENTRE

The Data Library is jointly operated by the UBC Computing Centre and the UBC Library. The cost of operation is jointly and quite fairly split between these two major service institutions, and very often, what one cannot get from one can be obtained from the other.

This special relationship, however, does mean that lines of responsibility and reporting are more than normally complex - ie. The Data Library does not report to one clearly defined superior officer, but at different levels, to different superiors in two discrete institutions.

1. Computing Centre

The programmer-analyst reports directly to the Senior Analyst, Numerical Analysis (at the time of writing, this is John Coulthard), and most relations with the Computing Centre are handled through this route - with some exceptions listed below.

Basically the unwritten agreement (as far as I understand it) is that the Computing Centre provides:

- a full-time programmer/analyst - and backup services when he/she is on vacation or requires assistance. This is a Professional and Administrative Staff position.
- CC$'s - at present about 5 CCIDs, and a seemingly unlimited amount of CC$s.
- necessary hardware, such as local terminals, etc.
- normal computing centre support provided to all users
- tape rack space
- general office supplies
- telephone, mail, photocopying
- physical space, renovations, repairs, and some of our furniture, thru Tom Taylor or Norman Wilson.

2. Library

The librarian reports technically through the Head of the Social Sciences Reference Division, but in actuality directly to the Assistant Librarian for Public Services. The Library assistant II is responsible, technically, to the Librarian, but in actual fact also to the programmer/analyst.
The Library provides to the Data Library:

- a full-time librarian
- a full-time library assistant II

- a real $ acquisitions budget
- cataloguing and other related support services
- acquisitions and processing services for the reference collection
- certain periodical subscriptions and other expensive reference items
- binding of monographs and periodicals
- supplies, stationary, etc.
- microfiching services
- general collections support

3. Other Departments And Faculties

Because the Data Library originally began in the Department of Political Science, and because this department continues to support the Data Library financially, the Department administers some of the benefits derived from the Data Library's membership in the Inter-University Consortium for Political and Social Research (ICPSR), ie the Department chooses the university's official representative to ICPSR, and chooses candidates for the ICPSR summer school courses.

B. PERSONNEL

1. Data Library personnel consists of three full-time positions:

- a full-time programmer/analyst who reports to the Senior Analyst, Numerical Analysis. This position is a 'professional and supervisory staff' position, and is governed by the policies of that union and the personnel policies of the Computing Centre. If the Comp Centre maintains a printed personnel policy I have never seen it; all personnel matters have in the past been handled through the Senior Analyst, including such things as time off, sick leave, vacations, etc. Vacations for this position are, however, coordinated with the other staff in the Data Library.

- the librarian reports technically through the Head of the Social Sciences Reference Division, but in actuality directly to the Assistant Librarian for Public Services. This positions is a Board of Governors appointment, and is governed by the professional personnel policies of the Library (see UBC Library policy and procedures manual). Records of sick leave, vacations, etc, have thus far been maintained by the secretary in the Social Sciences Reference Division. All personnel matters relating to this position should go through the assistant librarian responsible for professional staff. At some unspecified time in the past, and without notice, this position acquired the title Head, Data Library, but the Board of Governors still has it as a general librarian, and I have never pursued it.

- the Library assistant II is responsible, technically, to the Librarian, but in actual fact also to the programmer/analyst. This position is an AUCE governed position, and as such is governed by the personnel policies of the Library and by the AUCE contract. All matters relating to this position should go through the assistant librarian in charge of support staff (Erik de Bruijn) and the personnel secretary. Records of sick leave, vacations, etc. Are maintained by the Head, Data Library. Personnel records are to be found in the file folder 'personnel'; papers relating to AUCE, including a copy of the contract, are to be found in the file folder labelled 'AUCE'.

- occasionally, for special projects, student labour has been hired either part-time (10 hours/week) during the winter and spring terms, or full-time during the summer under various government sponsored employ-a-student schemes, usually through the library, but on occassion through the computing centre. Records are to be found in the file folder 'work study program'. Applications for such positions are invited at regular times of the year, usually by the Assistant Librarian for personnel (Erik de Bruijn).

2. General Policy Statement

For historical reasons as well as past experience, the following can be generalized about the problem of recruitment, background, and additional training of the three data library staff positions:

a. The Librarian

Should:
- have a social science background, for the time being. Many of the users come from the social sciences - those from other disciplines often are more sophisticated and do not require the level of assistance that the former require.
- have lots of reference experience - this job basically entails lots of reference know-how without a backup reference collection.
- a statistics background is helpful but not necessary, unless the programmer is weak in this area.
- be an outgoing personality, who can easily get along with people, and have a great deal of patience with repetitive, elementary questions. This is very much a public service position.

b. The Programmer

Should:
- have an academic degree in something! High-school graduation is simply not enough background when discussing research problems with faculty and graduate students - it basically boils down to a matter of maturity and breadth of general knowledge.
- a background in statistics, and/or social science or other quantitative research techniques is very helpful, especially if the librarian lacks these.
- have an extroverted personality, patience, and be able to talk to all kinds of people. This too is very much a public service position.

c. The Library Assistant II

Should:
- be hired from the library, preferably recruited from a division or position where he/she has had exposure to a wide variety of general library procedures, techniques, and policies. Experience has shown that it is much easier to teach someone familiar with library procedures the computer skills required in this job, than to try to teach someone with some computer skills such library skills as filing, main entry, acquisitions, binding, cataloguing, circulation, etc.
- for the same reasons as outlined above, this person too should be friendly, extroverted, and patient.

4. Training

- the librarian should without fail attend the ICPSR summer school course in mrdf management.
- and should also attend the ICPSR summer school in statistics.

The training schedule for a new staff member might look something like:

•• introduction to UBC Data Library functions - intensive
what are mrdf - different kinds
what does one do with them (research)
how are they stored, ie the tape collection
codebooks and other documentation, incl reference collection
read chapter 1 of Nasatir

•• tour of UBC Computing Centre facilities
4th floor offices - mail, photocopying, messages, etc
2nd floor terminal room, TEKTRONIX, printer, etc.
Basement information desk, printout, tape racks, etc.
read Introduction to UBC Computing Centre facilities

•• files (tape, and disc) & the file editor
read MTS files & devices

•• tape management - introduction
hanging tapes
TAPESHARE
internal tape labels
MGR
documentation, tapescans, partial dumps, etc
backups
archival tapes
read UBC tapeasy & UBC Tape

•• incoming mrdf
float versus permanent tapes - review
tapescan, MGR, etc - the FORMIN procedures

•• outgoing mrdf
release contract
MGR, etc - FORMOUT procedures

•• explanation of moral contracts
release contracts
user contracts

•• documentation - codebooks
various kinds of codebooks - substantive, physical formats
copying mrdf codebooks ($copy, osiris, spss, etc)
sending stuff to cataloguing - mrdf, refrence items
requesting spine labels
ordering reference material, and check in of same
serials check-in
transfers from other library locations
bindery
mendery

•• circulation
policies
circulation file
overdues and reminders

•• archival arrangements
archival tapes
Special Collections Division

•• user services
preliminary orientation
finding documentation
finding mrdf
SPIRES catalog
software help

•• the SPIRES databases

•• the reference collection, care and feeding
serials check-in
monographs, incomeing
cataloguing
binding
weeding
the card catalogue

C. BUDGET

The Data Library's budget is actually hidden in a number of different budgets, which are controlled by the Computing Centre and the Library.

The monies which pay the programmer/analyst are hidden in the Computing Centre's personnel budgets, and we need not concern ourselves with it.

The monies which pay the librarian and the library assistant II are contained (I understand) in a separate personnel budget, but as this is administered from the Librarian's office, we also need not concern ourselves with that.

The only budget with which it is necessary to concern ourselves is that part of the Library's acquisitions budget

that is assigned to Data Library acquisitions.

1. Data Library Acquisitions Account

The Data Library acquisitions account is numbered #1318, and is administered through the Accountant in the Acquisitions Division of the Library.

This account contains the real dollars used to pay all invoices, including:
- the purchase of tapes from the Computing Centre (done on a blanket requisition - see acquisition of tapes).
- payment of the ICPSR membership fee (annual), the Roper Center fee (annual), the COMPUSTAT subscription (annual), and the CRSP subscription (also annual).
- all other invoices for data files acquired, as they are sent to us.
- to this number are charged all reference items ordered in the normal way through Acquisitions Division (see the care and feeding of the reference collection).
- n.b. That the few serials in our reference collection are paid through the library's general serials accounts.

a. Getting

Every year in early spring (February?) the Library administration invites us to submit a projected budget submission for three years in the future. This budget submission should include all projected EXTRA costs for new programs, equipment, personnel, etc. General increases in the acquisitions budget are not included here. (see the file folder entitled Dlib budget).
Increases to the acquisitions budget, from one year to the next, are dependant on the total acquisitions budget the Library manages to get from the University administration.
Extra monies for the acquisition of additional expensive mrdf, such as the quinquennial Census data files, can be begged from the Collections Librarian (Tony Jeffries) out of the Research Books account.

b. Spending

The acquisitions budget must be spent, and all invoices submitted to the Accountant by the end of January (sometimes mid-February is ok) in order to be applied to the current budget year. Any monies not spent are lost, and cannot be carried over.

No arrangements have been made to encumber the account for mrdf that have been ordered, but for which no invoice has yet been received - however I am told that this can be arranged.

Once the funds are 'frozen' in February, in fact no invoices are paid until the new budget year commences April 1st, although invoices can be submitted, which will then be processed as soon as the new account monies have been assigned.

c. Kick-backs

As soon as the COMPUSTAT and CRSP invoices have been paid, and one knows how much was paid in Canadian dollars, a memo is sent to the Dean of the Faculty of Commerce requesting one-half that amount to be paid to the Data Library. (For copies of previous requests, see the file folder British Columbia. University. Faculty of Commerce.) N.B. That this should NOT include whatever is paid for the CRSP government bond files - the Library is picking up the full tab for this acquisition, since Commerce paid the full initiation fee.

As soon as the ICPSR fee has been paid and we have a total cost in Canadian dollars for that, a similar memo is sent to the Head of the Department of Political Science (copies of previous memos are to be found in the file folder 'British Columbia. University. Department of Political Science').

When the requisitions for these monies are received from the respective departments, they are photocopied for our own files, and the originals are sent to Accounts, Acquisitions with a request that Data Library account #1318 be credited with the appropriate amounts.

d. Processing Invoices

The steps in processing a standard invoice are as follows:

- check to make sure that the item(s) invoiced have in fact been received; if not, don't pay it.
- write on the invoice 'OK FOR PAYMENT DATA LIBRARY ACCOUNT #1318', stamp it with the date, and initial with the librarian's initials.
- take a special hollerith card (in file 'invoices for payment') stamp it with the same date as on the invoice, and initial it also with the librarian's initials.
- photocopy the invoice for the 'Dlib account' file.
- send the invoice, with the hollerith card clipped to it, and a covering note requesting that the invoice be paid from Data Library account #1318, to the Accountant, Acquisitions Div., Library Processing Centre.
- file the photocopy of the invoice in the 'Data Library account' file.
- when the monthly (?!?) statement of expenditures comes from Acquisitions, check to make sure that we

have only been charged for our own invoices, not for other items that we have neither ordered not authorized payment of.

2. Annual Reports, Etc.

a. Collecting Monthly Mount Statistics

Entering the command:

```
$SOURCE DLIB:UPDATE.$
```

causes the year-to-date file DLIB:Y_T_D to be updated by the contents of the logfile DLIB:MOUNTLOG, and produces listings of various logfiles including the tape-mount log, CHUM:OPLOG, DLIB:GRP.LOG, and the file -FILESTATUS (under id DLIB, ie all public files).

This should be done at the end of every month. The reports are filed in 2 3-ring binders marked 'MOUNTLOG' and 'OPLOG' behind the librarian's desk.

b. Printing A Year-end Mount Report

Entering the command:

```
$SOURCE DLIB:REPORT.$
```

prints a year-end summary report, including commercial versus academic tape mounts, a listing of all ids mounting Data Library owned tapes, and lists of tape files mounted during the past year in descending order of 'popularity' and by subject categories.

This should be run at the end of Aug 31st or Sept 1st each year to coincide with the beginning and end of the academic year. This is also filed in the MOUNTLOG binder behind the librarian's desk.

NOTE: that this is an ASAP routine.
ASAP writes to -P#. In order to get the report in 8x11 format, you have to:

```
$CON *PRINT* PORTRAIT ONESIDED
$COPY -P# TO *PRINT*
$REL *PRINT*
```

c. Others

- to the Library
Each year, when the summary report (see above) has been generated, a written report detailing the past year's accomplishments, and any upcoming general problems (as well as possible solutions thereto) is written for the Library, and sent to the Assistant Librarian for Public Serives (at time of writing, this is Bill Watson). A copy of this can also be sent to John Coulthard for further distribution within the Computing Centre.

- to IFDO
a copy of the same report, somewhat edited to eliminate any very specific local references, and concentrate on general activities and especially inter-archival cooperative efforts, is sent to all members of IFDO - this must be done, it is part of our membership contract with IFDO.

- to ICPSR, PAC/MRAD, and Essex
These three data suppliers (and possibly in future, others) have requested annual mount statistics. The easiest way to do this is to alter DLIB:REPORT.$ so that separate tables are produced for these suppliers with the catnames and any eventual mounts that may have occurred during the past year. However, this will require that all new data files acquired from these suppliers will have to be entered into the source.

3. Supplies, Equipment, Stationery, Etc.

Every February, all forms and supplies are ordered through the Library for the forthcoming year. The Library sends two separate forms, one for printed order forms, and the other for general supplies.

When ordering supplies, check to see what was ordered in previous years. To determine whether or not to re-order, count what is on hand and compare remaining supply with was ordered. Heavily used items like binders, accopress binders, folders, and labels are ordered every year. It is mostly common-sense.

When the order forms are filled out, a photocopy is kept in Dlib file 'supplies'. The 'printed order forms' order is sent to Gerry Dobbin in LPC, and the general order form is sent to the Librarian's Office.

During the following month supplies should start coming in, items order through the bookstore will be accompanied with a blue invoice. This invoice is mailed to Lily Wilson in the Librarian's Office. Note the date each item is received on your photocopy of the order form.

The order for printed forms functions much the same way.

In case supplies run very short, it is possible to 'borrow' from the Computing Center (4th floor office). When printed forms run low, contact Gerry Dobbin at LPC.

Pens, pencils, paper clips, etc, are also ordered from the Librarian's Office. One can either fill out a requisition form, (L-35 in supplies folder) or phone over the request. Depending on the size of order they will mail it or it can be picked up from the Librarian's office. 'Booster' orders can be submitted any time of the year. For items that can or need not be purchased right away, write a note in the supplies file so that they can be ordered the following February.

IV. A SHORT CHAPTER ON THE LONG SUBJECT OF CONTRACTS

Because mrdf are such an easily duplicated commodity, and because none of the standard concepts that exist to protect intellectual property can easily be applied to mrdf, an elaborate system of contracts and 'gentlemen's agreements have been built up in order to provide the sort of protection to data files as copyright legislation provides to published works.

Basically three kinds of contracts are at issue here:
- the contract between a principal investigator or producer of a data file, and a data archive or library in which the data file is to be deposited. This contract generally covers such issues as access and dissemination.
- the contract between a disseminating data archive or library and a recipient data archive or library or individual, to which an mrdf has been disseminated.
- the contract between a data archive or library and a user.

When an mrdf is disseminated from one data archive to another, an institution to a data archive, or an individual to a data archive, an implicit or explicit contract is made between the two parties. The disseminator is in fact binding himself/itself to 'stand behind' his data, and accept some or all of the responsibility for the physical and/or intellectual/ substantive quality of the data. The recipient is also entering into a contractual arrangement and it is these conditions that concern me here.

The primary concern of a release contract, ie contract covering the release of an mrdf to a data library or archive, quite apart from the question of reimbursement, revolve around access and dissemination. These conditions may be explicitly stated in a contract covering the transaction, or in a blanket contract covering all transactions between the depositor and the recipient, or it may be merely at the level of a gentlemen's agreement.

The question of access of course addresses the matter of WHO may use the mrdf in question, and under what circumstances. User populations often singled out for special mention include academic faculty, students and staff, ie the local academic community, adjacent academic communities, government, non-profit research institutions, and the 'commercial sector', ie those in the business of 'making money'. Restriction of access to any or most of these groups is normally explicitly stated. The conditions under which they are allowed access are very often left implicit, but are then normally universally understood to include acknowledgement of source if analysis results in publication, abrogation of responsibility on the part of principal investigator, producer, and disseminator for any interpretation or analyses of the data, and no illicit dissemination of copies of the data.

Dissemination is a rather separate problem, and in effect

consists of two issues: dissemination and redissemination. The distinction between the two can be most simply explained as follows: A produces a data file and disseminates a copy of it to B; can B then redisseminate copies of the data file to X, Y, and Z, and these in turn redisseminate copies to M, N, and P, or does the sole right of disseminating copies of this data reside with A? On the one hand, if A is an individual who does not wish to be further involved in dissemination of the mrdf, having, hopefully, once and for all deposited it in a data archive, he may be perfectly happy to allow his data to be freely disseminated and redisseminated, especially if he is concerned about easy access to data for verification of his research and secondary analysis of his data. On the other hand, if A is a data archive, which must justify its continued existence to its funding sources, it may very well restrict totally any redissemination of its data files, and maintain the sole right of redissemination. If redissemination is not restricted, other problems are generated, including quality and edition control both of the mrdf and of the documentation.

Generally, if no explicit contract covers the acquisition of a data file, I have made the following kinds of assumptions:
- if the UBC Data Library has paid money for the acquisition, and this money exceeds the cost of a tape, tape copying charges, and reasonable mail charges, then the mrdf is for the exclusive use of the UBC academic community, and is not to be redisseminated. Any requests for copies from outside the UBC academic community are referred to the original distributor.
- if we have acquired the data file for reasonable copying charges, or gratis, from an institution which is capable of redisseminating the mrdf, and likely to continue to do so in future, ie from an institution which is likely to continue to exist in future, then the data file can be used by anyone without restriction, but UBC Data Library will not redisseminate the data file, but will refer all requests for copies back to the original distributor, or ask the distributor for written permission to redisseminate.
- if the mrdf has been acquired for nominal or no cost from an individual or institution which is not likely to continue to exist or carry on its own dissemination of the data file, the mrdf can be accessed by anyone, and will be redisseminated by the UBC Data Library. But in this latter case, I will make every effort to make sure that we do get a signed release contract.

1. <u>Contracts For New Data Files Being Deposited</u>

A copy of our release contract covering data files being deposited in the Data Library follows.

When a data file is deposited, it should be signed in two copies, one to be kept by the depositor, and the other to stay in the Data Library. This latter copy is photocopied, and the original filed in Special Collections. The photocopy is stapled into the back of the file folder for the data file. And finally, a card is made up for the

'Dissemination file' including the principal investigator, the title, the access category, and (if access category A, or C, or F) the type of contract letter (see below).

2. Contracts For Data Files To Be Disseminated Elsewhere

The terms and conditions under which we disseminate data files to other institutions is determined by the conditions under which the data were deposited with us. If in doubt, consult the original release contract releasing the data to us. If we have no original release contract we probably cannot redisseminate the data file at all.

A library of FMT macros has been set up to generate the appropriate release contract for each data file. To find out which macro to use, look up the card for the data file in the 'Distribution file' - this will tell you to use LETTER1 or LETTER2, etc.

To any one of these macros must be fed three pieces of information for each output contract you wish to generate (ie each set of three generates one output contract):

- the title of the mrdf
- the name(s) of the principal investigator(s) in natural order,
- and the name of the recipient institution.

These MUST be in the above order, must be one line each (and ONLY one line each).

FMT feeds these into the appropriate slots in the output letter exactly as you have entered them.

The macrolibrary has been stored on the FS.0 tape, so to generate contracts, do the following:

```
$SIG CHUM PRIO=L COPIES=2 PORTRAIT
$$$$$$
$R DLIB:MOUNT
FS.0
$END
$R *FS 0=*T1*
R CONTRACT
STOP
$R *FMT
CALL LIBRARY(CHUM:CONTRACT)
CALL SETUP
Resource Manager
British Columbia political attitudes, 1978.
Daniel J. Koenig
University of Alberta
British Columbia political attitudes, 1979.
Daniel J. Koenig
University of Alberta
Quebec provincial election study, 1960.
```

Maurice Pinard, Albert Breton, and Raymond Breton
University of Alberta
$SIG

NB. You can only use one LETTER macro in each run, you cannot mix more than one kind of letter in any one run.

This will generate copies of the appropriate contracts ready for the intended recipient's signature.

3. Moral Contracts With Users

The issue here of course is to impress upon local users that they are not allowed to copy our data files for transportation to other institutions. There is not real way to police this effectively.

At one time we had all users sign a contract when they first came into the Data Library, agreeing not to make copies of files for dissemination purposes. This was found in the long run to be impossible to maintain.

Now, our continuing effort to impress this on users is to have an automatic message printed, whenever DLIB:MOUNT is run, to the effect that these files are not to be copied.

XIII. USER SERVICES

A. POLICY

This is the first time I have tried to write down any kind of service policy. Thus far, we have all sort of intuitively drawn the line at certain kinds of services, and bent over backwards to provide others. What I think it boils down to is this:

We will do everything possible to provide a user with a serious research project with the data he/she requires. This includes everything from acquiring the data, to helping the user read the codebook, consultation on the use of SPSS, OSIRIS, MVTAB, or any other statistical package, helping the user solve basic problems, even teaching them elementary MTS commands, and the use of the EDitor. Here it should be borne in mind that it is not our job to act as official Computing Centre consultants on any software other than that written specifically for the Data Library. Nor is it our job to act as consultants in such areas as statistical methodology, etc. But bear also in mind that not very much consulting is available elsewhere for undergraduate students, who are among our most numerous users - so we often end up teaching them basics.

What we will NOT do is do a user's work for him, including data analyses, etc. This is after all a learning institution.

B. CIRCULATION

Data library maintains a manual circulation system.
When a user wishes to borrow an item, a circulation card is filled out as indicated: especially important is the borrowers name, dept, and office telephone number (recorded in the space for address)
borrower's number is determined from library card
- stamp card and back of book with due date.
- initial card and file in circulation file.

1. Circulated Materials

- only the second (& third, fourth etc.) copies of codebooks are loaned. One copy must remain in the library.
- code books stamped "codebook only DLIB has no data" can be loaned.
- reference materials are not loaned for the most part.

2. Borrowers

- holders of AMS cards can borrow any codebooks
- persons who are not either UBC faculty, staff, or students are restricted as to the codebooks they can borrow.

CAN BE BORROWED:
- all codebooks describing MRDF in categories A & B
- codebooks in category E which have a copyright notice

CAN NOT BE BORROWED:
- codebooks in categories C & D
- codebooks in category E which do not have a copyright statement, often our own printouts of machine-readable codebooks

3. Date Due

- normally, circulation periods are determined on an ad hoc basis i.e. asking the user how long he/she will need the item; all within reason of course.
- if the codebook is part of a class assignment, the loan period would be greatly reduced, 1 day loan at most
- Canadian Census documentation is loaned for a period of 1 week only.

4. Overdues

- at certain intervals during the year, usually coincident with the Main librarys' call in notices, (this way the borrowers will be in the habit of returning books), the circulation file should be searched through for overdue loans.
- notices are sent to offending borrowers.
- if there is no address given, the Main Library circulation desk staff can supply it
- to some privileged borrowers (eg. Jean Laponce) a reminder only is sent once a year. For a decision on 'privileged borrowers', consult the librarian.
- the dates of all overdues and notices are stamped on the back of each card.
- if an item becomes outrageously overdue, and overdue notices are ignored, the matter is referred to Circulation Div (Joyce Harris) for handling.

5. Returns

When a borrowed item is returned, the appropriate circulation card is destroyed.

C. DECENTRALIZED CODEBOOK COLLECTIONS

In order to make some heavily used codebooks more accessible, we have put copies of them in other Library branches and divisions that have longer hours of opening. This includes:

COMPUSTAT	in Commerce Reading Room
CRSP user's guide	in Commerce Reading Room
CANSIM mini-base	in Social Sciences Div, Main Library in Government Publications Div.
CANSIM main base	in Social Sciences Div, Main Library in Government Publications Div.

No record of these except in my rather inaccurate memory - although there are circulation cards in the circ file for the items in Commerce Reading Room.

The CANSIM documentation is as far as I know not being used in Government Publications or Social Sciences Division, so this is rather wasted.

The CRSP and COMPUSTAT documentation in Commerce Reading Room should be updated as necessary as we get new CRSP and COMPUSTAT files.

D. DLIB:DOC

The purpose of this system was to supplement the UBC Data Library User's Guide, which is of course already out of date. The intent was to set up a system by which users could easily generate copies of any document they needed as a batch run, on their own, and not then need to actually come to the Computing Centre or the Data Library to get copies.

New documents are written as FMT source files. Both the FMT source file and the print-ready product of *FS are stored on the DOCFS1 tape (an FS tape, of course). The filename of the FMT source file should have '.S' as a suffix and that of the print-ready version should have '.W' as a suffix to distinguish them. When storing them on DOCFS1, use FS/DSAVE so that you can add a comment as to what is in the file.

DOCFS1 is backed-up on DOCFS2.

Paper copies of the documents are made - one for the documentation file in the Data Library user area, and one for the DLIB:DOC 3-ring binder.

Once you have a satisfactory document, it is time to add it to DLIB:DOC.

1. Maintenance Of DLIB:DOC

To add documents to those already available you should:
add to the file called SGA:DOCUMENTS.
add to the file called SGA:ALOWEDCMDS.
Source for the program DLIB:DOC is in SGA:DOCUMENT.

E. CLASSROOM USE

F. PUBLIC RELATIONS

G. INTRODUCTORY LECTURE

The following is a brief outline of a lecture given in October of 1981 as part of the Computing Centre's series of non-credit courses, in an attempt to introduce some new users to the services provided by the Data Library, to teach novices how to search with SPIRES, etc.

- Lecture: introduction to the Data Library.

- Outline

```
introduction                                    laine
  statement of objectives
    what we are
    what we do
    who we do it to
  what have we got                              Sophia
    SPIRES - searchable database
      EXPLAIN DATALIB
      EXPLAIN versus ..EXPLAIN
      search what?
        show searchterms
      search how?

        ..EXPLAIN SEARCHING
        ..EXPLAIN something
          FIND
          ALSO
          TYPE
          HELP
  other services                                laine
    provide documentation and access to mrdf
    acquisition of new mrdf
    consulting, basic handholding
    Data Library user's guide
    documentation system
      explanation
      menu of documents
```

editing batch run
what we do not do - runs, extensive programming
classroom use
prewarning of assignment
group instruction
individual hand-holding
prior idiot-proofing, system files, etc.
optimum codebook access
special purpose software sophia
GRP - for census data (Canadian) 1961-1976
financial databases, incl.

CRSP, a series of programs and subroutines
COMPUSTAT use DLIB:CPST (by Gladys Chiu)
CANSIM use DLIB:SDB
IMF data (IFS & DOT) use DLIB:IMF
depositing original mrdf laine
benefits of depositing
conditions of depositing

LIBRARY RESEARCH COMMITTEE
TEXAS A&M UNIVERSITY LIBRARY

RESEARCH PROPOSAL

Date: November 30, 1983

Name (principal investigator): Judith B. Droessler

Position: Reference Librarian Department: Reference

Co-investigator(s) if any (name, position): John Abbott, Reference Librarian

PROJECT TITLE: Nonbibliographic Databases in Texas: Availability and Use

Grant funds requested:

Travel	
Supplies	$110.00
Salaries	$236.60
Permanent equipment	
Computer time	
TOTAL:	$346.60

(For advice on expenditures, call Library Research Coordinator, 845-8111.)

BRIEF DESCRIPTION OF PROJECT, STATING PURPOSE, ANTICIPATED USES AND DISSEMINATION OF RESULTS OF PROPOSED PROJECT:

This research will explore two aspects of nonbibliographic databases: availability and use. Nonbibliographic databases generally hold data which are self-contained and are predominantly numeric and/or textual. Although bibliographic information may appear, the intent and composition of the database is numeric or textual. Some nonbibliographic databases are only available through purchase/lease of tapes, while others are accessible via online vendors. In many cases, access is expensive.

Faculty in selected departments at Texas A&M and other major Texas universities will be surveyed to determine which nonbibliographic databases are currently used, the amount of use, and types of users. The results of this inquiry will be of use to libraries and other institutions as an overview of the machine-readable nonbibliographic data files useful to researchers in specific disciplines. The Texas A&M portion of the survey will also provide an indication of the potential need for nonbibliographic data files at this university. A concurrent survey of academic research libraries in Texas will be conducted to determine on-site availability of nonbibliographic databases within the state. The results will be used to assess the current role of Texas university libraries in providing access to these files. In addition, the survey results will be used to compile a union list directory. This directory will be of use to subject researchers in locating needed tapes and to libraries, departments, and data processing centers in guiding acquisition of nonbibliographic machine-readable files. The results of the project will be disseminated through at least one article in a relevant publication (e.g. Texas Library Journal). The directory information may be submitted to the Texas State Library for publication.

On attached sheet provide the following information: brief statement of the problem, proposed methodology, objectives and goals, and anticipated results of this work. Provide a brief statement of the current state of the art relative to the propsoed work, supported by a bibliography of approximately 5-10 items.

STATE OF THE ART

Many non-bibliographic databases currently exist and they contain a wide range of political, historical, social, economic, scientific and other data. The formats, ease of use, cost, and accessibility of this information varies as widely as the contained subject matter. Although some non-bibliographic database directory sources do exist,[2] there is currently no single source which collates descriptive and/or location information on these data files.[3] In the absence of this information, researchers depend upon informal, often incomplete channels of referral to on-site tapes. There is also little information available regarding a reactive stance to data file acquisitions by libraries and other organizations.

There is currently disagreement as to the library's role in the provision of access to non-bibliographic databases. Rowe and Ryan[4] and Isaacson[5] argue convincingly that the library's responsibility as an information repository necessitates the acquisition of machine-readable data resources. Although a few libraries have successfully incorporated non-bibliographic files into their regular operations, libraries have generally been reluctant to acquire these databases.[6] Most machine-readable data files are currently administered by academic faculties, colleges, or departments.[7] As the number of machine-readable data files increases, library users may expect librarians to play an increasingly informed and proactive role in the dissemination of information from non-bibliographic databases.[8]

PROBLEM

The purpose of the proposed study is to explore the current role of Texas research libraries in relation to non-bibliographic databases, and to assess current usage of, and potential need for, these files among researchers in particular fields. Specific objectives of the research are: (1) To determine which databases are useful and in what intensity by researchers in Texas research universities, particularly Texas A&M University; and (2) To produce a union list form of directory of on-site non-bibliographic data files held by major universities in Texas.

METHODOLOGY

The local portion of the survey will be accomplished by personal interviews of faculty in selected departments conducted by a student worker hired for this project. Mailed questionnaires and telephone follow-up will be used for the state level part of the survey. A list of widely-used non-bibliographic databases will be compiled. Locally, targeted departments will be contacted and interviewed via faculty library representatives to pre-test a questionnaire and discover omitted aspects or databases. Both pre-test sample and total sample will be queried as to which databases they know to be currently available, and which databases they do or would make use of. Mailed surveys

(Please do not write below this line.)

Action of Committee: Date_____ Account Number_____

Recommended ( )
Not recommended ( )

_____
Library Research Committee

Amount_____

_____
Library Research Coordinator

_____
Library Director

will present a list of databases and ask for information regarding availability, awareness, and level of use. Although initial contacts will be made with library directors and department heads, the names of other appropriate contacts (such as data processing center personnel) will be solicited.

Data gathered by the surveys will be compiled in directory form with the assistance of a student employee hired for the project. The data will also be analyzed to determine patterns of nonbibliographic database usage in selected disciplines, and to assess the current state of library involvement in the provision of access to nonbibliographic files in Texas. This research is to also serve as a preliminary study for a future research project in which one or more nonbibliographic databases are offered on an experimental basis to Texas A&M University researchers to evaluate cost and usage factors.

REFERENCES

1. Byerly, Greg. Online Searching: A Dictionary and Bibliographic Guide. (Littleton, co: Libraries Unlimited, 1983).

2. Falk, Joyce Duncan, Sharon A. Hogan, and Judith S. Rowe. "Selected List of Guides to Databases and Data Archives." RQ, 21 (Summer 1982):363-364.

3. Rowe, Judith S. "Primary Data for Historical Research: New Machine-Researchable Resources." RQ, 21 (Summer 1982), 351-355.

4. Rowe, Judith and Mary Ryan. "Library Service From Numerical Data Bases: The 1970 Census as a Paradigm." College and Research Libraries, 35 (Jan. 1974), 7-15.

5. Isaacson, Kathy. "Machine-Readable Information in the Library." RQ, 22 (Winter 1982), 164-170.

6. Clark, Barton M. " Social Science Data Archives and Libraries: A View to the Future." Library Trends, 31 (Winter 1982), 505-509.

7. Ruus, Laine G.M. "Training of Data Services Professionals: Past, Present, and Future." Library Trends, 31 (Winter 1982):455-465.

8. Mignon, Edmond, " Numeric Data Bases in the Professional Librarianship Curriculum: Implications for Behavioral and Social Sciences Librarians." Behavioral and Social Sciences Librarian, 1 (Winter 1982), 181-187.

BUDGET DETAILS

TRAVEL: If you have requested travel funds for your proposed project, explain how the funds are to be spent.

Ground transportation: ____trips(s) to _____

totaling_____miles at _____cents per mile = $_____

Per diem: _____days at $_____ per day = $_____

Air transportation: _____trip(s) to _____=$_____

TOTAL TRAVEL FUNDS: $_____

SUPPLIES: Do you plan to use any sizable amount of materials such as card stock, carbon ribbons, bond paper, postage, or other consumable items? Even if you plan to use items usually available through the Library, you will need to estimate their costs so that the amount can be transferred from research to operating accounts. (Attached a list of any significant [i.e., expensive] items.)

TOTAL SUPPLIES: $ 110.00

SALARIES: Professional:____ _____month(s) at $_____

monthly gross salary = $_____

Clerical/student: 65 hours at $ 3.64 per hour =

$ 236.60

TOTAL SALARIES: $ 236.60

PERMANENT EQUIPMENT: Are there any items of equipment which need to be purchased in order to conduct this project?

TOTAL EQUIPMENT: $_____

COMPUTER TIME: Will you need to use the resources of the Data Processing Center to conduct your research? Will you need to make extensive use of AIRS (over and above the usual amount generally supported by the Library--e.g., compiling a voluminous, or potentially very expensive bibliography)?

DPC: Computing $_____ Programming Asst. $_____

AIRS Funds $_____

TOTAL COMPUTER: $_____

92 TOTAL FUNDS REQUESTED: $ 346.60

ICPSR

The University of Florida Libraries is a member of the Inter-University Consortium for Political and Social Research with the Social Sciences Librarian having a major responsibility for its functioning within the University and the State of Florida. The Social Sciences Librarian is the Organizational Representative of the ICPSR on campus. Moreover, the University of Florida Libraries pays the ICPSR fee each year. Members of ICPSR in Florida form a Florida Consortium. All requests for data sets from Florida State Universities and the University of Miami go through the University of Florida Libraries. The latter also serves as the ICPSR Archive for the state.

The Use of MRDFs in the Social Sciences: an anthropologist and a librarian look at the issues

H. Russell Bernard (Department of Anthropology)
and Ray Jones (Library)
University of Florida
Gainesville, FL 32611

The problems in making MRDFs accessible, as well as those in using them, are basically the same for all social scientists. Census tapes are more likely to be used by sociologists, and the Human Relations Area Files tapes are more likely to be used by anthropologists. But, from our perspective, these differences between disciplines are unimportant.

There are two kinds of problems: the quality of the MRDFs themselves and their accessibility. Quality control is the social scientists' headache; accessibility is the Librarian's.

First things first. No amount of computerizing data, or skill in massaging them, makes up for rotten data. Social researchers are trained to assess the reliability and validity of data. Putting data onto computer tapes and disks does not make those data any more valid than they would be if they remained on a room full of questionnaires. In the end, there is nothing that librarians can do to turn sow's-ear data into silk purses. That is the responsibility of the social scientists and social agencies that generate data in the first place. The librarian's role is not to assess the quality of MRDFs; it is to acquire such files and provide access to them.

Good, clean MRDFs, however, make it possible to ask questions that could not be asked in the past. Analyzing vast amounts of statistical data makes it possible to look for "weak signals" - potentially important relationships that do not emerge from sparse data. In any science, it is often the weak signals that are theoretically important. The complete results of surveys, censuses, and many longitudinal studies are available as MRDFs. Also small-area data can be retrieved from censuses. With the data in this form, operations can be performed, by scientists. Furthermore, MRDFs are not limited to the quantity of data provided, as are the statistical compendia or published tables. The 1980 census tapes contain approximately ten times more information than will be available in the published version. One can retrieve certain basic data at the city block level. With MRDFs the researcher can find any specific data points, create a sub-set and analyze them.

Microdata public use datasets offer an excellent example of the rich potential of MRDFs. For example, the 1980 Census Public Use Microdata consists of information about individual households and about the individuals comprising the household without any identification. Public use samples differ from the summary data which are already "summarized" prior to being formulated into tables. With the Census Public Use data, a subset can be retrieved; for example, for all persons who have more than $40,000 in income and who are older than 65. The demographic characteristics of this particular population can be statistically analyzed using programming packages. Moreover, there are 1940, 1950, 1960, 1970, and 1980 Census Public Use samples available. Other agencies (i.e., the Social Security Administration) also issue Public Use samples of their most significant data.

The Social Scientist now has available excellent statistical packages to

create customized files or subsets of major MRDFs. Packages can be used to analyze these files statistically, and to manage them.

The most widely used packages are SAS (Statistical Analysis System), SPSS (Statistical Package for the Social Sciences), and BMDP (Biomedical Data Program). CENSPAC, issued by the Census Bureau, has also become a standard package for manipulating census files. The statistical packages enable scientists to analyze, manipulate, and even create new variables. SAS, SPSS, and other packages are designed for non-programmers and are efficient for working with the many MRDF's produced by government agencies and with those surveys and polls resulting from individual or group studies.

Many social scientists have access to the vast data sources of the Inter-University Consortium for Political and Social Research (ICPSR). The ICPSR is one of the largest archives and disseminating services for MRDFs in the world. Begun in the early sixties, with emphasis on political science data, the Consortium now has a membership of over 260 colleges and universities that pay a yearly fee for access. The files cover all major social science areas including history, economics, general social surveys, and a number of interdisciplinary fields such as health services.

The datasets disseminated through ICPSR are formatted so that researchers may use standard statistical packages. For each dataset, there is a codebook describing its contents and the position and field where each data point may be located. The ICPSR has a representative on each campus of the subscriber universities. He or she is the OR (Organizaton Representative) and is the basic source for information about what is available from the Consortium, what files are on campus and what MRDFs may be requested. The Inter-University Consortium for Political & Social Research's Guide to Resources and Services 1982-83 is an excellent source for the librarian to learn of the scope and significance of MRDFs in the social sciences. The Guide is available from the OR, or from the Consortium at the University of Michigan.

The ICPSR does not currently give access to on-line numeric files or specific table-formatting software. It ships the datafiles and codebooks to the Organizational Representative of the member institution. These files are manipulated using the batch mode and statistical packages already mentioned.

There are, however, a number of on-line numerical datasets available to the public and software that will automatically produce tables.

Computers and statistical analysis have already affected research in political science, economics, business, and sociology, and is not altering anthropological research. The most important MRDF in anthropology is the on-line or tape versions of the Human Relations Area Files (HRAF). In 1975, HRAF produced the HRAFLIB (HRAF Hologeistic Computer Program Library). The HRAFLIB is a set of computer programs especially designed for specific what HRAF authors call "holocultural theory tests." These consist of correlational analyses of special characteristics of cultures in order to test hypotheses about cultural universals. Many of the statistical analysis performed by HRAFLIB are comparable to those performed by SAS and SPSS. A typical question put to the system might be the following:

> Is there a correlation between suicide and witchcraft
> in a sample of selected cultures?

The HRAFLIB manipulates anthropological data (traits) to test cross-cultural comparisons. The researcher would normally create subsets of traits and apply this software. HRAF also created some samples.

We stress, here, that librarians need not be familiar with the HRAF analysis software or with any of the statistical packages. These are tools of use to social scientists in analyzing data. In fact, there are real scientific dangers in having librarians (who are not usually well-trained in statistics) act as even the most low-level consultants on the use of SPSS and other similar packages. The basic rule is this: if you don't know your statistics, then computer packages won't help at all. One of us (Bernard) once had a student who exclaimed that he had found a "highly significant pattern of chi-squares in the data." The student had, in fact, used the wrong statistic to analyze the data. When asked what chi-square was, the student replied "Well, I don't really know, but it's what happened to my data when I told SPSS to run chi-square."

And this brings us to the real problem confronting librarians: making MRDFs accessible to social researchers.

Three factors affect accessibility: 1) Availability of bibliographic control. There is no integrated bibliographic control of MRDFs, no national Union Catalog of holdings. No accepted standards have yet been achieved for cataloging these materials. A researcher who wants to find books, periodicals, pamphlets or even broadsides on Colonial American political activity could expect to exhaust the resources of a major library through diligent use of its catalogs and bibliographies. The same researcher, seeking historical, statistical data in machine readable form would have to rely on luck (or running into a helpful, knowledgeable librarian who wasn't doing anything else at the moment) in order to exhaust the resources of a major MRDF holder.

2) Quality of Documentation. Even when an MRDF is located, it is not necessarily usable. Many MRDFs, like so many cleverly written computer programs, come with opaque documentation. Sometimes the miserable documentation is not even available. That is, a library may hold a tape, but the documentation lives at the source of the tape, and is available only on request. In the worst cases, no documentation is available. Quite often, universities hold data tapes at their computer centers or so-called "data labs," while documentation is held by the professor who brought the tape to campus.

This problem is widespread, and is indicative of the fuzzy-headed thinking about information storage and retrieval. Computers are so new to most administrators that they are easily bamboozled by folks who say: if it's a computer tape, then the computer center ought to be in charge of it. We are at a loss to explain this, since every computer center administrator we know admits that there are no empires to be built by gathering up all the tapes on campus. They are just another trivial management problem. With no intellectual reason to assign MRDF management to computer centers, and with no financial personnel or machine empires to built on them, there seems to be no explanation for the custom of keeping them at computer centers at all, rather than in libraries. But custom is not a trivial force in the world.

3) Level of user skills (or lack thereof). Quite simply, the majority of potential users of MRDFs in the social sciences can't. Graduate students don't learn about MRDFs, nor do they learn to access them because their professors have been left behind by the technology. We can depend on a continued increase in use

of MRDFs with no intervention on our part. But, as librarians, we can, and (we believe) we should intervene to teach the documentation skills (and not the statistical skills) that people need in order to use MRDFs.

Of course, special pedagogical skills are needed for this. That is, we need teachers (i.e., librarians) who know how to teach users (i.e., social scientists) how to turn on a computer and use it. We have noticed that many social scientists, including young graduate students as well as older scholars, go into something approaching an existential crisis when they approach computers and MRDFs for the first time. Interactive computer use is becoming part of the natural order for youngsters in school. But we have a responsibility, we believe, to teach working scholars about MRDFs. Scholars need instruction both on the content of MRDFs and on the methods for interrogating MRDFs.

If social scientists are, in general, ignorant about MRDFs, librarians are utterly innocent. Most of them have not even been exposed to numerical data sets. Furthermore, most librarians are like most of the public: they confuse computers with math and shun computers. There is a revolution going on with the sale of microcomputers to a vast, non-mathematical public. People everywhere are finding that computers are tools, like cars and telephones. No one needs to know anything about laser optics in order to make a trans-Atlantic call. And you don't need to know FORTRAN to do word-processing, or to do "on-line" bibliographic searches.

The social scientist does, indeed, have to know how to do statistical analysis of his or her data. But the librarian "only" needs to know how to a) scan the files to locate the correct MRDF; b) make it accessible to the end user. We put "only" in quotes because this is the weak link in the system, and the challenges for librarians are hardly trivial. By learning these new skills, librarians can take charge of a piece of the revolution, and not be swallowed by it.

In fact, social science Librarians are beginning to perceive the importance of machine-readable files. The Winter, 1982 issue of Library Trends was devoted to the challenges, problems, and models for setting up a data library either within or outside the academic library. Within the near future, social science librarians probably will not become data librarians, nor will many academic librarians form data centers of MRDFs. Instead, they will serve as "bibliographic brokers", i.e., while librarians will be working toward developing more effective access to MRDF, they will not actually possess the files, which will remain located in social science data archives."

A number of academic libraries in Florida and other parts of the United States have already placed the ICPSR codebooks in their reference collections, either cataloged or in a special area. More and more librarians have attended the workshop on data archives sponsored by the ICPSR. Hopefully, in the future, standard catalog entries for the machine-readable data files themselves can be utilized to form an OCLC data-base or a local database for library on-line catalogs. More and more graduate students and faculty in the social sciences are learning about certain important data files such as the General Social Survey, Census files, National Longitudinal Survey of Labor and Market Experience and Panel Study of Income Dynamics. Increasingly we find students and others asking us for codebooks of these datasets because fundamental studies have been made using these files. They are referenced as sources in their literature. Reference librarians who have some substantive knowledge of files are now asked such questions as "What sources should be consulted to study pension plans and their impact?" Knowing something of

the FASB II database and the Current Population Survey for May, 1979 can add to the patron's knowledge of potential sources. On a few occassions, the particular piece of data needed may be available only in a particular dataset.

Even without direct access to computer files or codebooks, the social science librarian should obtain printouts of major files that add to the use of data and to an understanding of its place in policy-making and research. Small-area data from the 1970 and 1980 census and a printout of the Equal Employment Opportunity Special File (EEO) are recommended. The latter is a primary source of data concerning racial and minority participation in over 500 occupational categories. Contact with a State Data Center or one of its affiliates can provide these printouts.

Librarians can meet their major responsibility in providing access to MRDFs by providing information about subject content and availability. As yet, there is no National Bibliography of MRDFs or central cataloging system for them. There are, however, several major catalogs from federal agencies, NTIS, and the ICPSR. The latter's Bulletin as well as the Newsletter of the Association of Public Data Users is very helpful. Another organization, the International Association for Social Science Information Services and Technology (IASSIST) publishes in the field. All the organization s disseminate information about MRDFs and some software systems. They have helped greatly to establish bibliographical standards for cataloging, but there is a lot left to be done.

Here is what librarians need to do right away:

1) Established librarians need to acquire new skills in the area known as data-base management. These are relatively easy skills to acquire, but in-service training is needed. There are very few libraries that provide that training for their staff. Requests by librarians will generate more programs.

2) As part of this training, librarians need to talk to systems programmers and other computer center personnel whose jargon is unfamiliar at first. Librarians need to learn how to set up a data archive and to back up MRDFs against damage and wipeouts. They must learn how to catalog data sets for users. Standards of bibliographic control will emerge as the cadre of MRDF bibliographers enlarges. Librarians need to take control of MRDFs, and to integrate them into the system of information storage and retrieval services already available in libraries. This will be welcomed by computer center personnel, if and only if librarians are clearly competent and can be trusted to provide proper care for MRDFs. Computer center personnel properly care a lot about proper care of MRDFs, but are not necessarily concerned about access. Librarians, whose career advancement may depend on developing greater access to some kind of information, need to show that they can be trusted to warehouse MRDFs before computer personnel will relinquish responsibility for those files.

3) Librarians need to develop acquisitions policies at their own institutions. We do not recommend standard acquisitions policies for all libraries. Each institution has its own perceived needs, and its own limited resources. Each institution must come to grips with several questions. What should one buy? Which MRDFs serve the library's constiuencies? How much should users pay, if anything, for access to MRDFs through libraries? There are different levels of service available. Should a library devote all its MRDF funds to a few on-line services, or to lots of less-expensive, but more cumbersome, batch services? These are management decisions and must be made by people who know enough about both MRDFs and their own user community to make such decisions intelligently.

4) Finally, librarians need to develop the same skills for MRDF consulting as they now have for consulting about print and microform materials. And they need to consider ways in which new technologies, now becoming available, can be of use. Telecommunications-computer links now make it possible for physicians in large hospitals to serve the diagnostic needs of rural colleagues. All they need is a terminal on their respective desks and an electronic mail system. Why shouldn't major libraries perform the same function with social data for those who need it at the local level? Librarians can take the lead in showing that computers can crunch words as well as numbers. And for those librarians who have the vision to develop such new services, there will be exciting opportunities and rewards.

SELECTED RESOURCES

Behavioral and Social Sciences Librarian. Vol. 3, No. 2, Winter 1983.

Chen, Ching-chih, ed. Numeric Databases, Norwood, N.J., Ablex, 1984.

Drexel Library Quarterly. Vol. 13, No. 1, January 1977.

Falk, Joyce Duncan, Sharon A. Hogan and Judith S. Rowe. Selected List of Guides to Databases and Data Archives. RQ Vol. 21, No. 4, pp. 363-364, Summer 1982.

Guide to Resources and Services. Ann Arbor, Mich., Inter-university Consortium for Political and Social Research, annual.

IASSIST Quarterly. Los Angeles, Calif., International Association for Social Science Information Services and Technology, quarterly.

ICPSR Bulletin. Ann Arbor, Mich., Inter-university Consortium for Political and Social Research, monthly.

Isaacson, Kathy. Machine-Readable Information in the Library. RQ Vol. 22, No. 2, pp. 164-170, Winter 1982.

Jones, Ray and Barbara Wittkopf. Computerized Census Data: Meeting Demands in an Academic Library. RQ Vol. 19, No. 3, pp. 246-251, Spring 1980.

Library Trends. Vol. 30, No. 3, Winter 1982.

McCrank, Lawrence J., ed. Automating the Archives, White Plains, N.Y., Knowledge Industry Publications, 1981.

Mignon, Edmond. Numeric Data Bases in the Professional Librarianship Curriculum: Implications for Behavioral and Social Sciences Librarians. Behavioral and Social Sciences Librarian, Vol. 1, No. 3, pp. 181-187, Spring 1980.

Rowe, Judith S. Primary Data for Historical Research: New Machine-Reasearchable Resources. RQ Vol. 21, No. 4, pp. 351-356, Summer 1982.